nestor

Referenzmodell für ein Offenes Archiv-Informations-System

- Deutsche Übersetzung -

nestor-Arbeitsgruppe OAIS-Übersetzung/Terminologie

nestor-materialien 16

Referenzmodell für ein Offenes Archiv-Informations-System

- Deutsche Übersetzung -

nestor-Arbeitsgruppe
OAIS-Übersetzung / Terminologie

nestor-materialien 16

Herausgegeben von

nestor - Kompetenznetzwerk Langzeitarchivierung und
Langzeitverfügbarkeit Digitaler Ressourcen für Deutschland

nestor - Network of Expertise in Long-Term Storage
of Digital Resources

http://www.langzeitarchivierung.de

Eine Bearbeitung der nestor-Arbeitsgruppe OAIS-Übersetzung / Terminologie:

- Georg Büchler (Koordinationsstelle für die dauerhafte Archivierung elektronischer Unterlagen, Schweiz)
- Niklaus Bütikofer (Historisches Institut der Universität Bern)
- Susanne Dobratz (HU Berlin)
- Katharina Ernst (Stadtarchiv Stuttgart)
- Christian Keitel (Landesarchiv Baden-Württemberg)
- Jens Ludwig (SUB Göttingen)
- Sabine Schrimpf (Deutsche Nationalbibliothek)
- Karin Schwarz (FH Potsdam)
- Sergio Gregorio (Schweizerische Bundeskanzlei, Zentrale Sprachdienste, Sektion Terminologie)

Basierend auf der Erstübersetzung des Historischen Archivs des Bayerischen Rundfunks:
Bettina Hasselbring, Brigitte Reimer, Christine Rost, Nicole Sachmann, Florian Westermayr

nestor-Arbeitsgruppe OAIS-Übersetzung / Terminologie
c/o Deutsche Nationalbibliothek
Sabine Schrimpf
Adickesallee 1
D-60322 Frankfurt am Main
Tel.: 069 / 15 25 17 61, Fax: 069 / 15 25 17 99
s.schrimpf@dnb.de

© 2012
nestor - Kompetenznetzwerk Langzeitarchivierung und Langzeitverfügbarkeit
Digitaler Ressourcen für Deutschland

Der Inhalt dieser Veröffentlichung darf vervielfältigt und verbreitet werden, sofern der Name des Rechteinhabers "nestor - Kompetenznetzwerk Langzeitarchivierung" genannt wird.
Eine kommerzielle Nutzung ist nur mit Zustimmung des Rechteinhabers zulässig.

URN: urn:nbn:de: 0008-2012051101

[http://nbn-resolving.de/urn/resolver.pl?urn=urn:nbn:de: 0008-2012051101]

Vorwort zur deutschen Übersetzung des OAIS-Referenzmodells

Das *Reference Model for an Open Archival Information System* (OAIS) ist sicherlich eins der einflussreichsten Dokumente im Bereich der Langzeiterhaltung digitaler Daten. Es beschreibt die Infrastruktur eines digitalen Langzeitarchivs, wobei es ausdrücklich über den technischen Aufbau des Archivs hinausgeht und die Organisation als Ganzes in den Blick nimmt. Es definiert zentrale Aufgabenbereiche und Verantwortlichkeiten des digitalen Langzeitarchivs, enthält Funktions- und Informationsmodelle und trifft wichtige terminologische Festlegungen. Da es neutral gegenüber Datentypen und –formaten, Systemarchitekturen und Institutionstypen ist, bietet das OAIS-Modell eine wesentliche Grundlage, über Spartengrenzen hinweg über digitale Langzeitarchivierung und digitale Langzeitarchive sprechen zu können.

Die Mitglieder der nestor-Arbeitsgruppe „OAIS-Übersetzung/Terminologie“ sind von der zentralen Bedeutung und dem konzeptionellen Nutzen des OAIS-Referenzmodells überzeugt. Mit der deutschen Übersetzung hoffen sie die Rezeption und Diskussion des Standards in der deutschen Fachgemeinschaft zu befördern. Sie erwarten nicht, dass die in der Übersetzung gewählten Begrifflichkeiten wie „Übergabeinformationspaket“ künftig in der deutschsprachigen Diskussion die eingeführte englische Kurzform „SIP“ ersetzt, oder von der „vorgesehenen Zielgruppe“ anstelle der „Designated Community“ gesprochen wird. Sie hoffen vielmehr, dass sich bei der Lektüre in der Muttersprache gerade die komplexeren Konzepte des OAIS-Modells leichter und schneller erschließen, und dass die Diskussion darüber anschließend umso besser geführt werden kann.

Das OAIS-Referenzmodell wurde 2002 vom *Consultative Committee of Space Data Systems* (CCSDS) als Empfehlung[1] und 2003 als ISO-Standard 14721:2003[2] veröffentlicht. Neben Vertreter führender Weltraumorganisationen waren an der Erarbeitung des Standards im CCSDS seit Mitte der 1990er Jahre auch Archiv- und Bibliotheksvertreter beteiligt. Nach einer 2006 eingeleiteten Review- und Kommentierungsphase liegt seit 2009 eine überarbeitete Version der CCSDS-Empfehlung vor (veröffentlicht als CCSDS Pink Book, August 2009[3]). Die Veröffentlichung dieser Version als ISO-Norm steht noch aus. Die vorliegende deutsche Übersetzung bezieht sich auf die überabeitete, frei verfügbare CCSDS-Version von 2009.

Die deutsche Übersetzung wurde von der nestor-Arbeitsgruppe „OAIS-Übersetzung/Terminologie“ auf Grundlage einer Erstübersetzung des Historischen Archivs des Bayerischen Rundfunks erarbeitet. Die Übersetzungsarbeit war von folgenden Prinzipien geleitet:

- Inhaltliche Korrektheit und sprachliche Präzision haben oberste Priorität.
- Das bedeutet aber auch, dass es der Genauigkeit halber an einigen Stellen zu hölzernen Formulierungen kommt. Die englische Originalvorlage bedient sich allerdings auch einer recht „technischen“ Sprache.
- Korrekte und einheitliche Verwendung der im Kapitel „Terminologie“ definierten Fachtermini, also wird z.B. „dissemination“ immer als „Auslieferung“ übersetzt. Liberalerer Umgang mit frei verwendeten Termini, so dass z.B. „set“ je nach Kontext als „Gruppe“, „Menge“ oder auch „Satz“ verwendet wird.

Der erste Überarbeitungsgang war vom Prinzip „inhaltliche Korrektheit und sprachliche Präzision“ geprägt. Alle Kapitel wurden von einem Erstbearbeiter durchgearbeitet und

[1] http://public.ccsds.org/publications/archive/650x0b1.pdf
[2] http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=24683. Während der ISO-Standard kostenpflichtig ist, ist die – inhaltlich identische – Version als PDF auf der Homepage des CCSDS frei verfügbar.
[3] http://public.ccsds.org/sites/cwe/rids/Lists/CCSDS%206500P11/Attachments/650x0p11.pdf

anschließend von einem Zweitbearbeiter überprüft. In einem dritten Überarbeitungsdurchgang wurden alle Kapitel auf die korrekte und einheitliche Verwendung der Fachtermini geprüft. Einige wenige übersetzte Stellen wurden von der Arbeitsgruppe mit Fußnoten versehen. Zum einen beziehen sich die Erläuterungen in diesen Fußnoten auf Fehler im englischen Original, zum anderen auf Begriffe, zu denen es keine Eins-zu-eins-Übersetzung gibt wie das anglo-amerikanische „Copyright“.

Leser, die in der Übersetzung trotz aller Gründlichkeit bei der Arbeit Fehler finden, werden gebeten, diese der Arbeitsgruppe mitzuteilen.

Der herzliche Dank der Arbeitsgruppe gilt den KollegInnen vom Historischen Archiv des Bayerischen Rundfunks, die die Erstübersetzung hergestellt haben, ohne die diese Arbeit gar nicht erst möglich gewesen wäre, und Herrn Alexander Zöller von der FH Potsdam, der die fertige Übersetzung einer gründlichen Endlektüre unterzog.

Vorwort[4]

Dieses Dokument ist eine technische Empfehlung zur Anwendung bei der Entwicklung eines breiteren Konsensus darüber, was Archive benötigen, um dauerhafte, oder unbegrenzt langfristige, Erhaltung digitaler Information zu gewährleisten.
Diese Empfehlung begründet einen gemeinsamen Rahmen von Begriffen und Konzepten, die ein Offenes Archiv-Informations-System (OAIS) ausmachen. Sie erlaubt einen aussagekräftigeren Vergleich und Gegenüberstellung bestehender und künftiger Archive. Sie stellt eine Grundlage für künftige Standardisierung im archivischen Kontext bereit, und sie sollte bei Herstellern zu einer größeren Kenntnis und Unterstützung archivischer Anforderungen führen.
Es wird erwartet, dass es durch den Prozess normaler Entwicklung an diesem Dokument zu Erweiterung, Streichung oder Modifizierung kommen kann. Dieser empfohlene Standard unterliegt daher den CCSDS Dokumentenmanagement- und Änderungssteuerungs-Verfahren, die im Verfahrenshandbuch für das Consultative Committee for Space Data Systems (*Procedures Manual for the Consultative Committee for Space Data Systems*) definiert sind.
Aktuelle Versionen von CCSDS-Dokumenten werden auf der CCSDS Website unterhalten: http://www.ccsds.org/
Fragen zu den Inhalten oder dem Status dieses Dokuments sollten an das CCSDS-Sekretariat gerichtet werden.

[4] Das originale CCSDS-Dokument enthält außer diesem Vorwort noch weitere, CCSDS-spezifische Vorbemerkungen, auf deren Übersetzung verzichtet wurde, da sie keine inhaltliche Relevanz für den Standard haben, sondern CCSDS-Gepflogenheiten würdigen: Autorität (*Authority*), Absichtserklärung (*Statement of Intent*), Liste der CCSDS-Mitglieder- und Beobachterorganisationen, Geleitwort (*Preface*) und Versionskontrolle (*Document Control*). Nach diesen Vorbemerkungen folgt das Inhaltsverzeichnis, die in der hier vorliegenden deutschen Übersetzung direkt an das Vorwort anschließt.

Inhalt

1 Einführung 2
1.1 Zweck und Aufgabenstellung 2
1.2 Anwendbarkeit 3
1.3 Begründung 3
1.4 Konformität 4
1.5 Leitlinien zur Entwicklung von verwandten Standards 4
1.6 Aufbau des Dokuments 6
1.7 Definitionen 7
2 OAIS-Konzepte 17
2.1. Die Umgebung des OAIS 18
2.2. OAIS Information 19
2.3 Außenbeziehungen des OAIS im Überblick 23
3 OAIS-Aufgabenbereiche 27
3.1 Verbindliche Aufgaben 27
3.2 Beispielhafte Mechanismen, wie man die Aufgaben erfüllen kann 27
4 Detaillierte Modelle 32
4.1. Funktionales Modell 32
4.2 Informationsmodell 50
4.3. Transformationen von Informationspaketen 77
5 Perspektiven der Erhaltung 82
5.1 Informationserhaltung 82
5.2 Erhaltung des Zugriffs- und Nutzungsservice 91
6 Interoperabilität von digitalen Archiven 95
6.1 Technische Ebenen der Kooperation von Archiven 95
6.2 Managementfragen bei Archivverbünden 101
Anlage zur deutschen Übersetzung 103

1 Einführung

1.1 Zweck und Aufgabenstellung

Der Zweck dieses Dokuments ist es, das Referenzmodell der Internationalen Organisation für Standardisierung (ISO) für ein **Offenes Archiv-Informations-System (OAIS)** zu definieren. Ein OAIS ist ein Archiv, das aus einer Organisation, die Teil einer größeren Organisation sein kann, aus Menschen und Systemen besteht, das die Verantwortung übernommen hat, Information zu erhalten und sie einer **vorgesehenen Zielgruppe** zugänglich zu machen. Es erfüllt eine Reihe von Verantwortlichkeiten, die in diesem Dokument definiert werden, und dies erlaubt die Unterscheidung von einem OAIS-Archiv von anderen Verwendungen des Begriffs „Archiv". Der Begriff „Offen" in OAIS impliziert, dass diese Empfehlung ebenso wie zukünftige Empfehlungen und Standards in offenen Foren entwickelt werden, nicht aber, dass der Zugriff auf das Archiv unbeschränkt ist.

Die zu erhaltende Information wurde als der Langzeitarchivierung bedürftig erachtet, auch wenn das OAIS selbst nicht von Dauer sein sollte. **Langfristig** bedeutet lange genug, um sich mit den Auswirkungen des Technologiewandels inklusive der Unterstützung neuer Datenträger und Datenformate oder mit einer wechselnden Benutzergruppe zu befassen. Langfristig kann sich auf unbestimmte Zeit erstrecken. In diesem Referenzmodell liegt ein besonderer Fokus auf digitaler Information, sowohl auf ursprünglichen Formen der vorgehaltenen Information als auch auf Zusatzinformation zu digital und physisch archiviertem Material. Daher bietet das Modell auch für solche Information Platz, die von Natur aus nicht-digital sind (z.B. eine *physische Probe*), die Modellierung und Erhaltung solcher Information wird aber nicht im Detail angesprochen. Dieses Referenzmodell:

- bietet einen Rahmen für das Verständnis und das gesteigerte Bewusstsein für archivische Konzepte, die zur langfristigen Erhaltung digitaler Information und den Zugriff auf sie nötig sind;

- bietet die Konzepte, die es nicht-archivischen Organisationen ermöglichen, erfolgreich am Erhaltungsprozess mitzuwirken;

- bietet einen Rahmen, inklusive Terminologie und Konzepte, zur Beschreibung und zum Vergleich des Aufbaus und der Tätigkeiten bestehender und zukünftiger Archive;

- bietet einen Rahmen zur Beschreibung und zum Vergleich verschiedener Langzeiterhaltungsstrategien und -techniken;

- bietet eine Basis zum Vergleich der Datenmodelle archivierter digitaler Information und für die Diskussion, inwiefern sich Datenmodelle und die zugrundeliegende Information über die Zeit verändern können;

- bietet einen Rahmen, der durch andere Bemühungen ausgeweitet werden kann, um auch die Langzeiterhaltung NICHT-digitaler Information (z.B. physische Datenträger oder physische Proben) abzudecken;

- erweitert den Konsens über die Elemente und Prozesse der digitalen Langzeiterhaltung und des langfristigen Zugriffs und begünstigt einen größeren Markt, den Anbieter unterstützen können;

- gibt Hinweise für die Identifikation und Anfertigung OAIS-verwandter Standards.

Das Referenzmodell spricht eine ganze Palette von Funktionen in der digitalen Langzeiterhaltung an, darunter Übernahme, Archivspeicher, Datenverwaltung, Zugriff und Auslieferung. Außerdem behandelt es die Migration digitaler Information auf neue Datenträger und in

neue Formate, die zur Repräsentation der Information verwendeten Datenmodelle, die Bedeutung der Software bei der Informationserhaltung sowie den Austausch digitaler Information zwischen Archiven. Es bestimmt sowohl interne als auch externe Schnittstellen zu den einzelnen Funktionen des Archivs sowie eine Reihe von übergeordneten Diensten zu diesen Schnittstellen. Es bietet diverse anschauliche Beispiele und einige „best practice"-Empfehlungen. Es definiert eine Mindestmenge von Verantwortlichkeiten für ein Archivs, damit es ein OAIS genannt werden kann, und es beschreibt ein Maximalarchiv durch breite Palette an nützlichen Begriffen und Konzepten.

1.2 Anwendbarkeit

Das OAIS-Modell in diesem Dokument kann in jedem Archiv angewandt werden. Es ist besonders geeignet für Einrichtungen, die die Verantwortung haben, Informationen langfristig verfügbar zu machen. Das schließt Einrichtungen mit anderen Zuständigkeiten ein, wie etwa die Sachbearbeitung und den Vertrieb als Reaktion auf programmatische Anforderungen.

Dieses Modell ist auch für solche Einrichtungen und Individuen von Interesse, die Information erzeugen, die langfristig erhalten werden sollen, und für solche, die Information von derartigen Archiven erhalten wollen.

Das Modell – inklusive der konzipierten Funktions- und Informationsmodelle – ist aus zwei Gründen relevant für den Vergleich und die Gestaltung von Einrichtungen, die Information für einen vorübergehenden Zeitraum bereithalten:

- Wenn man das rasante Tempo des technologischen Wandels oder mögliche Veränderungen in der vorgesehenen Zielgruppe bedenkt, werden aller Wahrscheinlichkeit nach Einrichtungen, die für einen vorübergehenden Zeitraum Information bereithalten, feststellen, dass einige oder viele der vorgehaltenen Informationen einer Beachtung der Langzeiterhaltung bedürfen.

- Auch wenn einige Einrichtungen, die Information vorhalten, selbst nur vorübergehend bestehen, können einige oder alle ihrer Informationen einer langfristigen Erhaltung bedürfen. Derartige Einrichtungen müssen sich unbedingt aktiv an den Bemühungen zur Langzeiterhaltung beteiligen.

Von Entwicklern von Standards wird erwartet, dass sie dieses Modell als Basis für künftige Standardisierung auf diesem Gebiet nutzen. Es ist eine Vielzahl von verwandten Standards denkbar. Ein Leitfaden für derartige Entwicklungen wird kurz in 1.5 angesprochen.

Dieses Referenzmodell gibt kein bestimmtes Design und auch keine bestimmte Art der Umsetzung vor. Tatsächliche Umsetzungen können Funktionen unterschiedlich gruppieren oder herausbrechen.

1.3 Begründung

Der enorme Anstieg der Rechnerleistung sowie der Bandbreiten und Verbindungen in Netzwerken bedingte eine explosionsartige Zunahme von Organisationen, die digitale Information zugänglich machen. Transaktionen zwischen allen möglichen Arten von Organisationen werden in digitaler Form ausgeführt, die an die Stelle traditionellerer Datenträger wie Papier treten.

Die Erhaltung von Information in digitaler Form gestaltet sich wesentlich schwieriger als die Erhaltung von Papier oder Film. Das stellt nicht nur traditionelle Archive vor Probleme, sondern auch viele Einrichtungen, die sich bisher gar nicht in einer Archivierungsfunktion gesehen haben. Die rapide Verbreitung digitaler Technologien in allen Bereichen bringt

gewisse Nachteile. Die Schnelllebigkeit der digitalen Technologien bringt beträchtliche technische Gefahren hervor, insbesondere ein viel größeres Risiko als in der Vergangenheit die Möglichkeiten der Restaurierung, Wiedergabe oder Interpretation einer Information zu verlieren. Wege, diese Gefahren zu vermeiden oder zu reduzieren werden durch diese Empfehlung detailliert. Nichtsdestotrotz wäre es unklug, dieses Problem allein vom technischen Standpunkt aus zu betrachten. Es gibt ebenfalls organisatorische, rechtliche, industrielle, wissenschaftliche und das Kulturerbe betreffende Aspekte zu berücksichtigen. Die Probleme zu ignorieren, die bei der Erhaltung von Information in digitaler Form entstehen, würde zwangsläufig zum Verlust dieser Information führen.

Es ist zu erwarten, dass dieses Referenzmodell, indem es Mindestanforderungen an ein OAIS-Archiv zusammen mit einer Reihe von archivischen Konzepten etabliert, einen allgemeinen Rahmen schafft, von dem aus die archivischen Herausforderungen, speziell in Bezug auf digitale Information, betrachtet werden können. Dies sollte mehr Organisationen befähigen die Anforderungen zu verstehen und entsprechende Maßnahmen zur Langzeiterhaltung von Informationen ergreifen. Darüber hinaus soll es eine Basis für mehr Standardisierung bieten und damit auch einen größeren Markt, auf dem Anbieter Anforderungen der Archive bedienen können.

1.4 Konformität

Eine OAIS-konforme Archiv-Umsetzung sollte das in 2.2 beschriebene Informationsmodell unterstützen. Das OAIS-Referenzmodell definiert oder verlangt dabei keine bestimmte Methode zur Umsetzung dieser Konzepte.

Ein OAIS-konformes Archiv soll den in 3.1 aufgelisteten Verantwortlichkeiten gerecht werden. Unterabschnitt 4 liefert Beispiele für die Mechanismen, mit deren Hilfe die in 3.1 identifizierten Verantwortlichkeiten erfüllt werden können. Diese Mechanismen werden für Konformität nicht verlangt. Es wird erwartet, dass ein separater Standard, wie in Abschnitt 1.5 dargestellt, erstellt wird, auf dem Akkreditierungs- und Zertifizierungsprozesse aufgebaut werden können.

Ein OAIS-konformes Archiv kann weitere Leistungen für Benutzer zur Verfügung stellen, die über die von einem OAIS verlangten hinausgehen.

Es wird angenommen, dass dieses Referenzmodell als Leitfaden dient, wenn eine spezifische Umsetzung entwickelt wird, um identifizierte Leistungen und Inhalte bereitzustellen. Dieses Dokument unterstellt oder befürwortet aber keine bestimmte Rechnerplattform, Systemumgebung, Systemarchitektur, Methode zur Systementwicklung, Datenbankmanagementsystem, Datenmodellierung, Datenbeschreibungssprache, Befehlssprache, Systemschnittstelle, Benutzeroberflächen, Technologie oder Datenträger, die für die Umsetzung erforderlich sind.

Das OAIS-Referenzmodell ist als konzeptioneller Rahmen entworfen worden, in dem über Archive gesprochen wird und in dem sie miteinander verglichen werden. Als solcher ist das Modell darum bemüht, alle wesentlichen Aktivitäten eines Langzeit-Archivs zu thematisieren und so einen einheitlichen und brauchbaren Satz an Begriffen und Konzepten zu definieren. Ein Standard und andere Dokumente, die den Anspruch erheben konform zum OAIS-Referenzmodell zu sein, sollten die im OAIS-Referenzmodell definierten Begriffe und Konzepte in der gleichen Weise verwenden.

1.5 Leitlinien zur Entwicklung von verwandten Standards

Dieses Referenzmodell, das vom CCSDS Gremium 2 als Antwort auf ISO TC20/SC 13 entwickelt wurde, dient dazu, weitere Gebiete zu bestimmten, für die OAIS-verwandte Standards entwickeln werden können. Einige dieser Standards werden vielleicht vom

CCSDS entwickelt; andere wiederum von anderen Standardisierungs-Gremien. Wenn diese Arbeit von anderen Gremien aufgenommen wird, sollte sie jedoch dermaßen koordiniert werden, dass Unverträglichkeiten und Anstrengungen möglichst gering gehalten werden. Einige Bereiche für mögliche OAIS-verwandte Standards sind unten mit Angaben zu relevanten Standards und/oder Initiativen aufgelistet:

- Standard(s) für die Schnittstellen zwischen OAIS-konformen Archiven;

- Standard(s) für die in einem Archiv verwendete Übergabe (Übernahme)-Methodik ;

  - ISO 20652:2006 Space data and information systems - Producer-Archive Interface Methodology Abstract Standard

- Standard(s) für die Übergabe (Übernahme) digitaler Datenquellen ins Archiv;

- Standard(s) für die Lieferung von digitalen Quellen aus dem Archiv;

- Standard(s) für die Übergabe von digitalen Metadaten über digitale oder physische Datenquellen ins Archiv;
  - PREMIS Data Dictionary for Preservation Metadata. Version 2.0, PREMIS, Editorial Committee, March 2008;
  - ISO 15889:2003 Space data and information transfer systems—Data Description Language—EAST Specification;
  - ISO 21961:2003 Space data and information transfer systems—Data Entity Dictionary Specification Language (DEDSL)—Abstract syntax;
  - ISO 21962:2003 Space data and information transfer systems—Data Entity Dictionary Specification Language (DEDSL)—PVL syntax;
  - ISO 22643:2003 Space data and information transfer systems—Data Entity Dictionary Specification Language (DEDSL)—XML/DTD;

- Syntax-Standard(s) für die Identifizierung von digitalen Quellen innerhalb des Archivs;

- Protokoll-Standard(s) für die Suche und das Auffinden von Metadateninformationen über digitale und analoge Datenquellen;

- Standard(s) für den Zugriff auf Datenträger, die es erlauben, Datenträgerverwaltungssysteme auszutauschen, ohne dass die Datenträger neu
beschrieben werden müssen

- Standard(s) für bestimmte physische Datenträger;

- Standard(s) für die Migration einer Information auf andere Datenträger oder
in andere Formate;

- Standard(s) für empfohlene Vorgehensweisen bei der Archivierung;
  - ISO 15489-1:2001 Information and documentation—Records management. Part 1: General;
  - ISO/TR 15489-2:2001 Information and documentation—Records management.
  - Part 2: Guidelines;
  - ISO 23081-1:2006 Information and documentation—Records management
  - processes—Metadata for records—Part 1: Principles;
  - ISO/TS 23081-2:2007 Information and documentation—Records management
  - processes—Metadata for records—Part 2: Conceptual and implementation issues;

- Standard(s) für die Zertifizierung von Archiven
  - Trustworthy Repositories Audit & Certification: Criteria and Checklist (TRAC), Version 1.0 (February 2007).

1.6 Aufbau des Dokuments

1.6.1 Wie sie dieses Dokument lesen

Alle Leser sollten die Unterabschnitte „Zweck und Aufgabenstellung“ (1.1), „Anwendbarkeit“ (1.2) und „Konformität“ (1.4) lesen, um einen Überblick über die Themen und Aufgabenstellungen dieses Dokuments zu gewinnen.

Diejenigen, die lediglich einen Überblick über die Hauptkonzepte wollen, sollten außerdem „OAIS-Konzepte“ (Abschnitt 2) und „OAIS-Aufgabenbereiche“ (Abschnitt 3) lesen.

Wer ein OAIS-Archiv umsetzen oder tagtäglich verwalten möchte, sollte das ganze Dokument lesen.

1.6.2. Organisation der Abschnitte

Abschnitt 1 liefert Zweck, Aufgabenstellung und Anwendbarkeit des Modells sowie Definitionsabschnitte, die für viele Standards typisch sind. Außerdem liefert es eine Begründung für die Bemühung, Konformitätsanforderungen und einen Leitfaden für die Entwicklung verwandter Standards.

Abschnitt 2 bietet einen Überblick auf hoher Ebene über die Hauptkonzepte, die in ein OAIS-Archiv eingebunden sind. Es vermittelt einen Blick auf die Umgebung eines OAIS-Archivs und auf die Rollen derjenigen, die in Interaktion mit ihm stehen. Es erörtert, was mit dem Begriff „Information“ gemeint ist und was notwendig ist, um Information langfristig zu erhalten. Es beinhaltet wesentliche Konzepte, die wichtig sind für OAIS-konforme Realisierungen.

Unterabschnitt 3.1. definiert verbindliche Aufgaben, die ein OAIS-Archiv erfüllen muss, um seine Information zu archivieren, und 3.2. enthält erläuterndes Material über die Instrumente, die viele Archive brauchen können, um diese Aufgaben zu erfüllen.

Abschnitt 4 enthält Modellansichten, die man braucht, um ein OAIS-Archiv detaillierter zu verstehen. Es bricht das OAIS herunter auf eine Anzahl verschiedener Funktionsbereiche und identifiziert einige übergeordneten Dienste zu den Schnittstellen. Es enthält ebenfalls detaillierte Datenmodellansichten von Information in Form von UML-Diagrammen.

Abschnitt 5 bietet einige Perspektiven auf die Probleme bei der Informationserhaltung durch Migration auf andere Datenträger, in andere Formate oder andere Repräsentationen. Ebenso liefert es einige Perspektiven auf die Frage nach der Erhaltung von Zugriffsdiensten auf digitale Information, etwa durch Software-Porting, Software-Einbettung oder die Emulation von Hardware.

Abschnitt 6 ist eine Einführung in die verschiedenen Möglichkeiten für die Zusammenarbeit von Archiven, um bessere oder kosteneffektivere Diensten anbieten zu können.

Die Anhänge B-E sind nicht Teil der Empfehlung und werden für den Leser als zusätzlicher Service bereitgestellt.[5]

- Anhang A bietet ein zusammengesetztes Diagramm der in 4.1 beschriebenen detaillierten Funktionsbereiche

[5] Die Anhänge sind in der deutschen Übersetzung nicht enthalten.

- Anhang B verknüpft Teile dieses Referenzmodells mit anderen Standardwerken;

- Anhang C bietet eine kurze Einführung in die Unified Modeling Language (UML);

- Anhang D enthält eine Liste mit Literaturangaben;

- Anhang E liefert ein Schichtenmodell der Information;

1.7 Definitionen

1.7.1 Akronyme und Abkürzungen[6]

AIC	Archival Information Collection – *Archivinformationskollektion*
AIP	Archival Information Package – *Archivinformationspaket*
AIU	Archival Information Unit – *Archivinformationseinheit*
ASCII	American Standard Code for Information Interchange
CCSDS	Consultative Committee for Space Data Systems
CD-ROM	Compact Disk - Read Only Memory – Compact Disk
CRC	Cyclical Redundancy Check
DBMS	Data Base Management System
DDL	Data Description Language
DIP	Dissemination Information Package – *Auslieferungsinformationspaket*
DRM	Digital Rights Management
FITS	Flexible Image Transfer System
HFMS	Hierarchical File Management System
IEEE	Institute of Electrical and Electronic Engineers
ISBN	International Standard Book Number
ISO	International Organization for Standardization
NASA	National Aeronautics and Space Administration
OAIS	Open Archival Information System – *Offenes Archiv-Informations-System*
PDI	Preservation Description Information – *Erhaltungsmetadaten*
SIP	Submission Information Package – *Übergabeinformationspaket*
UML	Unified Modeling Language
UNICODE	Universal Code
WWW	World Wide Web

1.7.2. Terminologie

In diesem Referenzmodell werden viele Begriffe gebraucht, deren Bedeutungen sehr genau definiert werden müssen. Diese Begriffe werden in diesem Unterabschnitt definiert. Bei ihrer erstmaligen Nennung im Text werden sie fett und großgeschrieben dargestellt. Danach werden sie nur noch großgeschrieben.[7]

[6] Hier werden lediglich deutsche Übersetzungen der Glossar-Begriffe gegeben. Auf Übersetzung der nicht im OAIS-Glossar enthaltenen Fachbegriffe wird verzichtet. Im folgenden Übersetzungstext werden die teilweise auch in der deutschen Fachsprache sehr gebräuchlichen englischen Akronyme beibehalten. SIP steht dann z.B. für *Auslieferungsinformationspaket.*

[7] In der deutschen Schriftsprache funktioniert die Hervorhebung von Begriffen durch Großschreibung nicht in der Weise, in der sie im Englischen funktioniert. In der deutschen Übersetzung wurde daher auf weitere Hervorhebungen der Begriffe, nachdem sie bei der Erstnennung fett gesetzt wurden, verzichtet. Einzige Ausnahme: In Kap. 4 „Detaillierte Modelle“ sind die Namen der Funktionseinheiten in ‚einfache Anführungszeichen’ gesetzt, um die Lesbarkeit zu unterstützen.

Weil dieses Referenzmodell in allen möglichen Disziplinen und Einrichtungen anwendbar ist, die digitale Information erhalten und bereitstellen oder dies bald tun wollen, können diese Begriffe nicht mit allen Begriffen übereinstimmen, die sonst in diesen Disziplinen geläufig sind, z.B. in traditionellen Archiven, digitalen Bibliotheken oder Forschungsdatenzentren. Stattdessen ist der Ansatz dieser, Begriffe zu verwenden, die nicht schon mit Bedeutung überladen sind, um so die Gefahr zu reduzieren, nicht beabsichtigte Bedeutungen zu vermitteln. Daher wird erwartet, dass alle Disziplinen und Einrichtungen feststellen werden, dass sie einige ihrer gewohnteren Begriffe auf die des OAIS-Referenzmodells übertragen müssen. Das sollte nicht schwierig sein und wird eher als Beitrag zum Erfolg des Referenzmodells gesehen denn als Abschreckung. Zum Beispiel zielt die Archivwissenschaft auf die Erhaltung der „Akte". Dieser Begriff wird im OAIS-Referenzmodell nicht verwendet, aber man könnte ihn ungefähr gleichsetzen mit „Inhaltsinformation innerhalb eines Archivinformationspakets" (siehe die folgenden Definitionen sowie 2.2 und 4.2 für den Zusammenhang). Einige wenige solcher Übertragungsbeispiele werden in Anhang B aufgeführt.[8]

Abgeleitetes AIP (Derived AIP)
Ein AIP, generiert durch Extraktion oder Ansammlung von Information aus einem oder mehreren Quell-AIPs.

Allgemeine Dienste (Common Services)
Die unterstützenden Dienste wie Interprozesskommunikation, Namensdienste, kurzfristige Speicherzuteilung, Ausnahmebehandlung, Sicherheit und Verzeichnisdienste, die nötig sind um das OAIS zu unterstützen.

Allgemeine Nutzergruppe (Global Community)
Eine erweiterte Endnutzergruppe im Kontext eines Archivverbunds, die auf Bestände von verschiedenen Archiven über ein oder mehrere gemeinsame Findmittel zugreifen kann.

Archiv (Archive)
Eine Einrichtung, die beabsichtigt, Information für den Zugriff und die Verwendung durch eine vorgesehene Zielgruppe zu erhalten.

Archiveinheits-Beschreibung (Unit Description)
Eine Art der Paketbeschreibung, die speziell Information über eine Archivinformationseinheit zur Verwendung durch die Zugriffshilfen bereitstellt.

Archivinformationseinheit (AIU) (Archival Information Unit (AIU))
Ein Archivinformationspaket, bei dem das Archiv entscheidet, die Inhaltsinformation nicht in weitere Archivinformationspakete aufzuschlüsseln. Eine AIU kann aus mehreren digitalen Objekten bestehen (z.B. mehrere Dateien).

Archivinformationskollektion (AIC) (Archival Information Collection (AIC)
Ein Archivinformationspaket, dessen Inhaltsinformation sich aus verschiedenen anderen Archivinformationspaketen zusammensetzen.

Archivinformationspaket (AIP) (Archival Information Package (AIP))
Ein Informationspaket, bestehend aus der Inhaltsinformation und den dazugehörigen Erhaltungsmetadaten, das innerhalb eines OAIS aufbewahrt wird.

Archivverbund (Federated Archives)
Eine Gruppe von Archiven, die sich darauf verständigt hat, Zugriff auf ihre Bestände über eine oder mehrere gemeinsame Findmittel zu ermöglichen

[8] Anhang B ist in der deutschen Übersetzung nicht enthalten.

Auffrischung (Refreshment)
Eine digitale Migration, bei der ein Datenträger durch eine Kopie ersetzt wird, die hinlänglich exakt ist, das die gesamte Hard- und Software des Archivspeichers weiter läuft wie zuvor.

Auslieferungsinformationspaket (DIP) (Dissemination Information Package (DIP))
Ein Informationspaket, abgeleitet aus einem oder mehreren AIPs, das der Endnutzer als Antwort auf eine Anfrage an das OAIS erhält.

Authentizität (Authenticity)
Das Ausmaß, in dem eine Person (oder System) ein Objekt als das ansieht, was es vorgibt zu sein. Authentizität wird auf der Basis von Evidenz beurteilt.

Bearbeitetes AIP (AIP Edition)
Ein AIP, dessen Inhaltsinformation oder Erhaltungsmetadaten angereichert oder verbessert wurden, nicht mit der Absicht, Information zu erhalten, sondern um es zu verstärken oder zu verbessern. Ein bearbeitetes AIP wird nicht als Ergebnis einer Migration betrachtet.

Begleitbeschreibung (Associated Description)
Die Information, die den Inhalt eines Informationspakets aus Sicht einer bestimmten Zugriffshilfe beschreibt

Bereitstellungssystem (Retrieval Aid)
Eine Anwendung, die es berechtigten Benutzern ermöglicht, die durch die Paketbeschreibung beschriebene Inhaltsinformation und Erhaltungsmetadaten abzurufen.

Bestellsystem (Ordering Aid)
Eine Anwendung, die den Endnutzer bei der Ermittlung der Kosten von, und der Bestellung von, gewünschten Archivinformationspakete unterstützt.

Bestellvereinbarung (Order Agreement)
Eine Vereinbarung zwischen dem Archiv und dem Endnutzer, in der die physischen Details über die Lieferung, wie etwa Art des Datenträgers und Datenformate, festgelegt sind.

Daten (Data)
Eine in formalisierter Weise rückinterpretierbare Repräsentation von Information, die zur Kommunikation, Interpretation oder Verarbeitung geeignet ist. Beispiele für Daten beinhalten eine Bitsequenz, eine Zahlentabelle, die Buchstaben auf einer Seite, die Tonaufnahmen einer sprechenden Person oder eine Mondgesteinsprobe.

Datenauslieferungssitzung (Data Dissemination Session)
Eine Lieferung von Datenträgern oder eine einzelne Telekommunikations-Sitzung, die Daten einem Endnutzer Daten zur Verfügung stellt. Das Format/die Inhalte der Datenauslieferungssitzung basieren auf einem Datenmodell, das in der Bestellvereinbarung zwischen dem OAIS und dem Endnutzer ausgehandelt wurde. Dieses Datenmodell identifiziert die logischen Konstrukte, die das OAIS verwendet, und wie sie bei jeder Datenträger-Lieferung oder in der Telekommunikations-Sitzung repräsentiert werden.

Daten der Datenverwaltung (Data Management Data)
Die Daten, die im permanenten Speicher der Datenverwaltung erzeugt und gespeichert werden und sich auf den Archivbetrieb beziehen. Einige Beispiele für diese Daten sind Buchhaltungsdaten für Endnutzerrechnungen und -berechtigungen, Policy-Daten, Daten über Ereignisbasierte Bestellungen (Abonnements) für sich wiederholende Anfragen, Daten über die Entwicklung von Erhaltungsprozessen sowie statistische Daten zur Erstellung von Berichten für das Archivmanagement.

Daten-Lexikon (Data Dictionary)
Ein formaler Speicher an Begriffen zur Beschreibung von Daten.

Datenobjekt (Data Object)
Entweder ein physisches Objekt oder ein digitales Objekt.

Datenübergabesitzung (Data Submission Session)
Eine Lieferung von Datenträgern oder eine einzelne Telekommunikations-Sitzung, die dem OAIS Daten zur Verfügung stellt. Das Format/die Inhalte der Datenübergabesitzung basieren auf einem Datenmodell, das in der Übergabevereinbarung zwischen dem OAIS und dem Produzenten ausgehandelt wurde. Dieses Datenmodell identifiziert die logischen Konstrukte, die der Produzent verwendet, und wie sie bei jeder Lieferung von Datenträgern oder in einer Telekommunikations-Sitzung repräsentiert werden.

Digitale Migration (Digital Migration)
Der Transfer digitaler Information innerhalb des OAIS mit dem Ziel ihrer Erhaltung. Sie unterscheidet sich von Transfer im Allgemeinen in drei Punkten:

- einem Fokus auf der Erhaltung des gesamten Informationsgehalts, der Erhaltung benötigt;
- einer Perspektive, dass die neue archivische Erscheinung der Information ein Ersatz für die alte ist; und
- dem Verständnis, dass die volle Kontrolle und Verantwortung über alle Aspekte des Transfers bei dem OAIS liegen.

Digitales Objekt (Digital Object)
Ein Objekt, das sich aus einer Reihe von Bitsequenzen zusammensetzt.

Endnutzer (Consumer)
Die Rolle, die von Personen oder Client-Systemen eingenommen wird, die mit Diensten des OAIS interagieren, um erhaltene Information zu finden und im Einzelnen auf diese Information zugreifen zu können. Das kann andere OAIS sowie OAIS-interne Personen oder Systeme einschließen.

Ereignisbasierte Bestellung (Event Based Order)
Eine Endnutzeranfrage nach Information, welche regelmäßig mit Eintreten eines oder mehrerer Ereignisse geliefert werden sollen.

Erhaltungsmetadaten (PDI) (Preservation Description Information (PDI))
Die Information, die benötigt wird, um die Inhaltsinformation angemessen zu erhalten, und die als Provenienz, Referenz, Beständigkeit, Kontext und Information über Zugriffsrechte kategorisiert werden kann.

Erschließungsinformationen (Descriptive Information)
Der Satz an Informationen, der vor allem aus Paketbeschreibungen besteht und der der Datenverwaltung zur Verfügung gestellt wird, um das Finden, Bestellen und Abrufen des Informationsbestands eines OAIS durch Endnutzer zu unterstützen.

Findmittel (Finding Aid)
Eine Art der Zugriffshilfe, die es dem Benutzer ermöglicht, gewünschte Archivinformationspakete zu suchen und zu identifizieren.

Funktionseinheit Administration (Administration Functional Entity)
Die OAIS-Funktionseinheit, die die Dienste und Funktionen enthält, die gebraucht werden, um tagesaktuell den Betrieb der anderen OAIS-Funktionseinheiten zu überprüfen.

Funktionseinheit Archivspeicher (Archival Storage Functional Entity)
Die OAIS-Funktionseinheit, die die Dienste und Funktionen enthält, die für die Speicherung und das Wiederauffinden der Archivinformationspakete gebraucht werden.

Funktionseinheit Datenverwaltung (Data Management Functional Entity)
Die OAIS-Funktionseinheit, die die Dienste und Funktionen zur Bestückung, Aufrechterhaltung und Zugreifen auf eine große Vielfalt an Informationen umfasst. Einige Beispiele für solche Informationen sind Kataloge und Inventare dessen, was vom Archivspeicher abgerufen werden kann, verarbeitende Algorithmen, die über die abgerufenen Daten laufen können, Endnutzer-Zugriffs-Statistiken, Endnutzerrechnungen, Ereignisbasierte Bestellungen, Sicherheitskontrollen sowie Ablaufpläne, Policies und Verfahren des OAIS.

Funktionseinheit Erhaltungsplanung (Preservation Planning Functional Entity)
Die OAIS-Funktionseinheit, die die Dienste und Funktionen zur Überwachung der Umgebung des OAIS enthält und Empfehlungen und Erhaltungspläne zur Verfügung stellt, um sicherzustellen, dass die Information, die im OAIS gespeichert ist, für die vorgesehene Zielgruppe langfristig zugänglich und verstehbar bleibt, auch wenn die ursprüngliche Rechenumgebung obsolet wird.

Funktionseinheit Übernahme (Ingest Functional Entity)
Die OAIS-Funktionseinheit, die die Dienste und Funktionen zur Annahme von Übergabeinformationspaketen von Produzenten enthält, die Archivinformationspakete für die Speicherung vorbereitet und sicherstellt, dass Archivinformationspakete und ihre unterstützenden Erschließungsinformationen innerhalb des OAIS etabliert werden.

Funktionseinheit Zugriff (Access Functional Entity)
Die OAIS-Funktionseinheit, die die Dienste und Funktionen enthält, welche die Archivinformationsbestände und damit zusammenhängende Dienste für die Endnutzer sichtbar machen.

Grundwissen (Knowledge Base)
Eine Reihe von Informationen, verinnerlicht in einer Person oder einem System, die es der Person oder dem System erlauben, empfangene Information zu verstehen.

Inhaltsdatenobjekt (Content Data Object)
Das Datenobjekt, das zusammen mit der Repräsentationsinformation die Inhaltsinformation enthält.

Inhaltsinformation (Content Information)
Ein Satz an Informationen, der das eigentliche Ziel der Erhaltung ist oder der Teile der oder die komplette Information enthält. Es ist ein Informationsobjekt, das sich aus dem Inhaltsdatenobjekt und seiner Repräsentationsinformation zusammensetzt.

Information (Information)
Jede Art von Wissen, das ausgetauscht werden kann. Während des Austauschs wird es durch Daten repräsentiert. Ein Beispiel wäre eine Bitfolge (die Daten), begleitet von einer Beschreibung, wie die Bitfolge als Zahlen zu interpretieren ist, die eine Temperaturmessung in Celsius (die Repräsentationsinformation) darstellen.

Information über Zugriffsrechte (Access Rights Information)
Die Information, die die Inhaltsinformation betreffende Zugriffsbeschränkungen identifiziert, einschließlich der rechtlichen Rahmenbedingungen, Lizenzbedingungen, und Zugriffskontrolle. Sie beinhaltet die Zugriffs- und Verbreitungsbedingungen, die in der Übergabevereinbarung angegeben wurden, bezogen sowohl auf die Erhaltung (durch das OAIS) und die letztendliche Nutzung (durch den Endnutzer). Sie beinhaltet außerdem die Spezifikationen für die Anwendung von Maßnahmen zur Durchsetzung von Rechten.

Informationseigenschaft (Information Property)
Der Teil der Inhaltsinformation, der von der Informationseigenschaftsbeschreibung beschrieben wird. Der detaillierte Ausdruck, oder Wert, dieses Teils des Informationsinhalts wird von den geeigneten Teilen des Inhaltsdatenobjekts und seiner Repräsentationsinformation übermittelt.

Informationseigenschaftsbeschreibung (Information Property Description)
Die Informationseigenschaftsbeschreibung. Es ist eine Beschreibung eines Teils des Informationsinhalts eines Inhaltsinformationsobjekts, der für einen bestimmten Zweck hervorgehoben wird.

Informationsobjekt (Information Object)
Ein Datenobjekt zusammen mit seiner Repräsentationsinformation.

Informationspaket (Information Package)
Ein konzeptioneller Container, der sich aus optionaler Inhaltsinformation und optional dazugehörigen Erhaltungsmetadaten zusammensetzt. Zu diesem Informationspaket gehört Verpackungsinformation, welche die Inhaltsinformation und die Paketbeschreibungsinformation, die Suchen nach der Inhaltsinformation ermöglicht, voneinander abgrenzt und identifiziert.

Irreversible Transformation (Non-Reversible Transformation)
Eine Transformation, für welche nicht garantiert werden kann, dass sie eine reversible Transformation ist.

Kollektionsbeschreibung (Collection Description)
Eine spezielle Art von Paketbeschreibung, die Information über eine Archivinformationskollektion für Zugriffshilfen bereitstellt.

Kontextinformation (Context Information)
Die Information, die die Beziehung der Inhaltsinformation zu ihrer Umgebung wiedergibt. Dazu gehört, warum die Inhaltsinformation erzeugt wurde und wie sie sich zu anderen Inhaltsinformationsobjekten verhält.

Kooperierende Archive (Co-operating Archives)
Diejenigen Archive, die vorgesehene Zielgruppen mit ähnlichen Interessen haben. Sie können untereinander Daten bestellen und übernehmen. Kooperierende Archive müssen sich einigen, mindestens ein gemeinsames Übergabeinformationspaket und Auslieferungsinformationspaket für Anfragen zwischen den Archiven unterstützen.

Langfristig (Long Term)
Eine Zeitspanne, die lange genug andauert, um sich mit den Auswirkungen des Technologiewandels inklusive der Unterstützung von neuen Datenträgern und Datenformaten sowie einer sich verändernden vorgesehenen Zielgruppe auf die Information im OAIS auseinander zu setzen. Diese Zeitspanne reicht bis in die unbestimmte Zukunft.

Langzeiterhaltung (Long Term Preservation)
Die langfristige Erhaltung von Information in einer für die vorgesehene Zielgruppe unmittelbar verstehbarer Form, und mit Evidenznachweisen, die ihre Authentizität langfristig unterstützten.

Lokale Nutzergruppe (Local Community)
Die ursprüngliche, von einem Archiv bediente, vorgesehene Zielgruppe, im Kontext eines Archivverbunds.

Management (Management)
Die Rolle, die von denjenigen eingenommen wird, die die OAIS-Policy als einen Bestandteil in einem umfassenderen Policy-Bereich etablieren, zum Beispiel als Teil einer größeren Organisation.

Metadaten (Metadata)
Daten über andere Daten.

Nachfolgeplan (Succession Plan)
Der Plan, wie und wann Management, Besitz und/oder Kontrolle an den OAIS-Beständen einem nachfolgenden OIAS übergeben werden, um die kontinuierliche, effektive Erhaltung dieser Bestände zu gewährleisten.

Offenes Archiv-Informations-System (OAIS) (Open Archival Information System (OAIS))
Ein Archiv, das aus einer Organisation, die Teil einer größeren Organisation sein kann, aus Menschen und Systemen besteht, das die Verantwortung übernommen hat, Information zu erhalten und sie einer vorgesehenen Zielgruppe zugänglich zu machen. Es erfüllt eine Reihe von Verantwortlichkeiten, wie in Abschnitt 4 definiert, die es erlauben, ein OAIS-Archiv von anderen Verwendungen des Begriffs „Archiv“ zu unterscheiden. Der Begriff „Offen“ in OAIS soll andeuten, dass diese Empfehlung ebenso wie zukünftige, verwandte Empfehlungen und Standards in offenen Foren entwickelt werden, und nicht, dass der Zugriff auf das Archiv unbeschränkt ist.

Paketbeschreibung (Package Description)
Die Information, die zur Verwendung durch Zugriffshilfen vorgesehen ist.

Persistenzinformation (Fixity Information)
Die Information, welche die Mechanismen dokumentiert, die sicherstellen, dass das Inhaltsinformationsobjekt nicht unerlaubt verändert wurde. Ein Beispiel ist ein Schlüssel aus einer zyklischen Redundanzüberprüfung (CRC) für eine Datei.

Physisches Objekt (Physical Object)
Ein Objekt (so wie Mondgestein, ein Bio-Präparat, ein Objektträger) mit physikalisch wahrnehmbaren Eigenschaften, die Information repräsentieren, die für geeignet gehalten werden, angemessen für die Erhaltung, Verbreitung und eigenständige Nutzung dokumentiert zu werden.

Produzent (Producer)
Die Rolle, die von denjenigen Personen oder Client-Systemen eingenommen wird, die die Information zur Verfügung stellen, die erhalten werden soll. Dies kann andere OAIS oder OAIS-interne Personen oder Systeme einschließen.

Provenienzinformation (Provenance Information)
Die Information, welche die Geschichte der Inhaltsinformation dokumentiert. Diese Information teilt den Ursprung oder die Quelle der Inhaltsinformation mit, jegliche Änderungen seit ihrer Erzeugung und wer sie seit ihrer Entstehung in Obhut hatte. Das Archiv ist verantwortlich dafür, Provenienzinformation ab dem Zeitpunkt der Übernahme zu erzeugen und zu erhalten; frühere Provenienzinformation sollte allerdings vom Produzenten zur Verfügung gestellt werden. Provenienzinformation ergänzt die Evidenznachweise, die Authentizität unterstützen.

Recherchesitzung (Search Session)
Eine vom Endnutzer initiierte Sitzung mit dem Archiv, während der der Endnutzer die Findmittel des Archivs benutzen wird, um möglicherweise interessante Bestände zu identifizieren und zu untersuchen.

Referenzinformation (Reference Information)
Die Information, die als Identifikator für die Inhaltsinformation verwendet wird. Sie beinhaltet auch Identifikatoren, die es externen Systemen erlauben, bestimmte Inhaltsinformation eindeutig zu referenzieren. Ein Beispiel für eine Referenzinformation ist eine ISBN.

Referenzmodell (Reference Model)
Ein Bezugssystem zum Verständnis von signifikanten Beziehungen zwischen den Einheiten der Umgebung sowie zur Entwicklung von einheitlichen Standards oder Spezifikationen zur Unterstützung dieser Umgebung. Ein Referenzmodell basiert auf einer kleinen Anzahl vereinheitlichender Konzepte und kann als Grundlage verwendet werden, um Laien zu schulen und ihnen Standards zu erklären.

Replikation (Replication)
Eine digitale Migration, bei der es keine Änderungen an der Verpackungsinformation, den Inhaltsinformation und den Erhaltungsmetadaten gibt. Die Bits zur Repräsentation dieser Informationsobjekte werden bei dem Transfer auf denselben oder einen neuen Datenträgertyp erhalten.

Repräsentationsinformation (Representation Information)
Die Information, die ein Datenobjekt in für Menschen aussagekräftigere Konzepte übersetzt. Ein Beispiel von Repräsentationsinformation für eine Bitsequenz, die eine FITS-Datei ist, könnte sich aus einem FITS-Standard, der das Format definiert und einem Wörterbuch, das die Bedeutung von Schlüsselbegriffen definiert, die nicht Bestandteil des Standards sind, zusammensetzen.

Repräsentationsnetzwerk (Representation Network)
Der Satz von Repräsentationsinformation, der die Bedeutung eines Datenobjekts vollständig beschreibt. Repräsentationsinformation, die in digitalen Formen vorliegt, braucht ergänzende Repräsentationsinformation, damit deren digitale Formen langfristig verstanden werden können.

Repräsentations-Wiedergabe-Programm (Representation Rendering Software)
Eine Art von Software, die die Repräsentationsinformation eines Informationsobjekts in für Menschen verstehbaren Formen darstellt.

Reversible Transformation (Reversible Transformation)
Eine Transformation, in der die neue Repräsentation eine Menge (oder eine Unter-Menge) resultierender Einheiten definiert, die den von der ursprünglichen Repräsentation definierten resultierenden Einheiten entsprechen. Das bedeutet, dass es eine Eins-zu-eins-Übersetzung zurück zur ursprünglichen Repräsentation und ihrer Menge von Ausgangs-Einheiten gibt.

Semantische Information (Semantic Information)
Die Repräsentationsinformation, die weitergehend die Bedeutung, über die von der Strukturinformation vermittelte hinausgehend, beschreibt.

Spontan-Bestellung (Adhoc Order)
Eine Endnutzer-Anfrage nach Information, für die das OAIS angezeigt hat, sie sei zurzeit verfügbar.

Strukturinformation (Structure Information)
Die Repräsentationsinformation, die erklärt, wie andere Information organisiert ist. Zum Beispiel übersetzt sie Bitstreams in geläufige Computer-Zeichen wie Buchstaben, Zahlen und Pixel sowie in Gruppierungen solcher Zeichen wie Buchstabenketten und -reihen.

Teilbeschreibung (Member Description)
Eine Begleitbeschreibung, die einen Teil einer Kollektion beschreibt.

Transformation (Transformation)
Eine digitale Migration, bei der es eine Änderung an der Inhaltsinformation oder den Erhaltungsmetadaten eines Archivinformationspakets gibt. Zum Beispiel ist die Änderung von ASCII Codes in UNICODE in einem Textdokument, das erhalten wird, eine Transformation.

Transformationelle Informationseigenschaft (Transformational Information Property)
Eine Informationseigenschaft, deren Erhaltung als notwendig, aber nicht ausreichend angesehen wird, um nachzuweisen, dass eine irreversible Transformation den Informationsinhalt angemessen erhalten hat. Dies könnte wichtig sein, weil es zu den Evidenznachweisen über die Authentizität beiträgt. Solche Informationseigenschaften werden mit bestimmter Repräsentationsinformation verknüpft werden müssen, einschließlich der semantischen Information, um zu verzeichnen, wie sie codiert sind und was sie bedeuten. (Der Begriff „Significant Property“, der in der Literatur unterschiedlich definiert wird, wird manchmal so verwendet, dass er mit dem der transformationellen Informationseigenschaft, übereinstimmt.)

Transformiertes AIP (AIP Version)
Ein AIP, dessen Inhaltsinformation oder Erhaltungsmetadaten eine Transformation auf das Quell-AIPs durchlaufen hat, und das als Ersatz für das Quell-AIP vorgesehen ist. Ein transformiertes AIP wird als Ergebnis einer digitalen Migration betrachtet.

Übergabeinformationspaket (SIP) (Submission Information Package (SIP))
Ein Informationspaket, das vom Produzenten an das OAIS geliefert wird, um es zur Konstruktion oder zur Aktualisierung eines oder mehrerer AIPs und/oder den dazugehörigen Erschließungsinformationen zu benutzen.

Übergabevereinbarung (Submission Agreement)
Die Vereinbarung zwischen dem OAIS und dem Produzenten, die ein Datenmodell spezifiziert und jegliche anderen benötigten Absprechen für die Datenübergabesitzung. Dieses Datenmodell bestimmt Format/Inhalte und die logischen Konstrukte, die der Produzent verwendet, und wie sie bei jeder Lieferung eines Datenträgers oder in einer Telekommunikations-Sitzung repräsentiert werden.

Übersichtsbeschreibung (Overview Description)
Ein Spezialfall der Kollektionsbeschreibung, die eine Kollektion in ihrer Gesamtheit beschreibt.

Umverpackung (Repacking)
Eine digitale Migration, bei der es einige Veränderungen an der Verpackungsinformation des Archivinformationspakets gibt.

Unmittelbar verstehbar (Independently Understandable)
Ein Merkmal von Information, die hinreichend vollständig ist, um von der vorgesehenen Zielgruppe interpretiert, verstanden und verwendet zu werden, ohne dass diese auf spezielle, nicht weit verbreitete, Hilfsmittel, einschließlich benannter Personen, zurückgreifen muss.

Verpackungsinformation (Packaging Information)
Die Information, die verwendet wird, um die einzelnen Bestandteile eines Informationspaketes zusammenzuhalten und zu identifizieren. Es könnten zum Beispiel die ISO-9660 „Volume Deskriptoren“[9] und ein Dateiverzeichnis sein, die auf einer CD-ROM verwendet werden, um den Inhalt mehrerer Dateien, die die Inhaltsinformation und die Erhaltungsmetadaten enthalten, zur Verfügung zu stellen.

[9] In OAIS steht hier nur ISO 9660 volume – es macht aber nur volume descriptors Sinn.

Vorgesehene Zielgruppe (Designated Community)
Eine ausgewiesene Gruppe potenzieller Endnutzer, die in der Lage sein sollen, einen bestimmten Satz an Informationen zu verstehen. Die vorgesehene Zielgruppe kann sich aus mehreren Benutzergruppen zusammensetzen. Eine vorgesehene Zielgruppe wird vom Archiv definiert und diese Definition kann sich über die Zeit verändern.

Zugriffshilfe (Access Aid)
Ein Software-Programm oder Dokument, das es Endnutzern ermöglicht, die gesuchten Archivinformationspakete zu finden, auszuwerten und abzurufen.

Zugriffskollektion (Access Collection)
Eine Kollektion von AIPs, die durch eine Kollektionsbeschreibung beschrieben wird, für die aber keine Verpackungsinformation im Archivspeicher existiert.

Zugriffsprogramm (Access Software)
Eine Art von Software, die Teile des oder den gesamten Informationsgehalt eines Informationsobjekts in für Menschen oder Systeme verstehbaren Formen präsentiert.

2 OAIS-Konzepte

Der Zweck dieses Abschnitts ist es, einige zentrale OAIS-Konzepte zu begründen und zu beschreiben. Eine ausführlichere Beschreibung sowie eine formale Modellierung dieser Konzepte erfolgt in Abschnitt 4.

Der Begriff „Archiv“ bezieht sich mittlerweile auf eine große Bandbreite von Funktionen und Systemen zur Speicherung und Erhaltung. Klassische Archive werden als Einrichtungen oder Organisationen verstanden, die ursprünglich durch oder für eine Organisation, Institution oder Körperschaft der Regierung und für den Zugriff durch öffentliche oder private Gruppen eingerichtet wurden. Das Archiv erfüllt diese Aufgabe, indem es die Unterlagen in seinen Besitz übernimmt, indem es gewährleistet, dass die Unterlagen für die darauf zugreifende Gruppe verstehbar bleiben und indem es die Unterlagen so behandelt, dass ihr Informationsgehalt und ihre Authentizität erhalten bleiben. Geschichtlich gesehen bestanden diese Unterlagen aus Büchern, Papieren, Karten, Fotografien und Filmen, welche von Menschen unmittelbar oder mithilfe einfacher optischer Vergrößerungs- und Abtasthilfen gelesen werden können. Der Hauptfokus bei der Erhaltung dieser Information lag darin, sicherzustellen, dass sie sich auf Medien mit einer langfristigen Haltbarkeit befinden und dass der Zugriff auf diese Medien sorgfältig kontrolliert wird.

Das explosionsartige Wachstum von Information in digitaler Form stellt nicht nur die klassischen Archive und ihre Informationszulieferer, sondern auch viele Organisationen der Regierung, der Privatwirtschaft und des Non-Profit-Sektors vor eine massive Herausforderung. Diese Organisationen erkennen oder werden bald erkennen, dass sie Funktionen zur Informationserhaltung übernehmen müssen, die typischerweise mit klassischen Archiven assoziiert werden, weil digitale Information leicht verloren geht oder korrumpiert wird. Das Tempo der technologischen Entwicklung bringt es mit sich, dass einige Hardware und Software-Systeme innerhalb weniger Jahre obsolet werden. Dieser Wandel kann großen Druck auf die Eignung der entsprechenden Datenstrukturen oder Formate ausüben, weiterhin diese Information wie gewünscht repräsentieren zu können. Weil viele der Zusatzinformationen, die für den Erhalt dieser Information notwendig sind, leichter verfügbar oder sogar nur zu der Zeit verfügbar sind, in der die Originalinformation hergestellt wird, müssen sich diese Organisationen aktiv an den Bemühungen zur Langzeiterhaltung beteiligen. Dabei ist es nötig, dass sie den Grundsätzen dieses OAIS-Referenzmodells folgen, um sicherzustellen, dass die Information langfristig erhalten werden kann. Die Beteiligung an diesen Bemühungen wird die Lebenszyklus-Kosten minimieren und eine erfolgreiche Langzeiterhaltung der Information ermöglichen.

Die Explosion der Rechnerleistung und der digitalen Datenträger hat sich oft in Systemen ereignet, in denen die Produzentenrolle und die Archivrolle in die Zuständigkeit der gleichen Stelle fallen. Diese Systeme sollten sich auf die in diesem Dokument diskutierten Ziele der Langzeiterhaltung verpflichten. Der Gestaltungsprozess muss erkennen, dass einige Langzeiterhaltungsaktivitäten mit den Zielen einer schnellen Produktion und raschen Auslieferung von Produkten an Endnutzer in Konflikt geraten können. Die Gestalter und Architekten solcher Systeme sollten die erreichten Lösungen dokumentieren.

Ein Hauptziel dieses Referenzmodells besteht darin, ein größeres Verständnis dafür zu vermitteln, was für eine langfristige Erhaltung von und den Zugriff auf Information erforderlich ist. Um Verwechslungen mit einfachen Bit-Speicher-Funktionen zu vermeiden, definiert das Referenzmodell ein Offenes Archiv-Informations-System (OAIS), welches eine Funktion zum langfristigen Erhalt von Information und zum Zugriffe ausübt.

Ein OAIS ist ein Archiv, das Information für eine vorgesehene Zielgruppe zugreifbar und nutzbar halten will und das die in Abschnitt 6[10] aufgeführten Anforderungen erfüllt. Gemeint sind sowohl Archive, die kontinuierliche Informationsströme entgegennehmen müssen, als auch Archive, die vorwiegend unregelmäßige Zugänge erfahren. Es schließt sowohl Archive ein, die eine Vielzahl hochentwickelter Zugriffsdienste anbieten, als auch solche Archive, die nur einfachste Anfragen unterstützen. Für den Rest dieses Dokuments werden die Begriffe „Archiv“ und „OAIS“ gleichgesetzt und jeweils auf ein OAIS-Archiv bezogen, es sei denn der Kontext legt ein anderes Verständnis nahe (z.B. klassisches Archiv).

Das OAIS-Modell erkennt an, dass digitale Informationsbestände bereits jetzt stark dezentralisiert sind, und es weiß um den Bedarf an lokalen Umsetzungen von effizienten Policies und Prozeduren zur Unterstützung der Informationserhaltung. Dies erlaubt beim Umsetzen grundsätzlich eine große Bandbreite von organisatorischen Lösungen, die auch verschiedene Rollen für klassische Archive mit einschließen. Voraussichtlich werden Organisationen beim Versuch, Information zu erhalten, erkennen, dass OAIS-Begriffe und Konzepte ihnen bei der Erreichung ihrer Erhaltungsziele helfen.

2.1. Die Umgebung des OAIS

Das einfache Modell in Abbildung 2-1 zeigt die Umgebung eines OAIS.

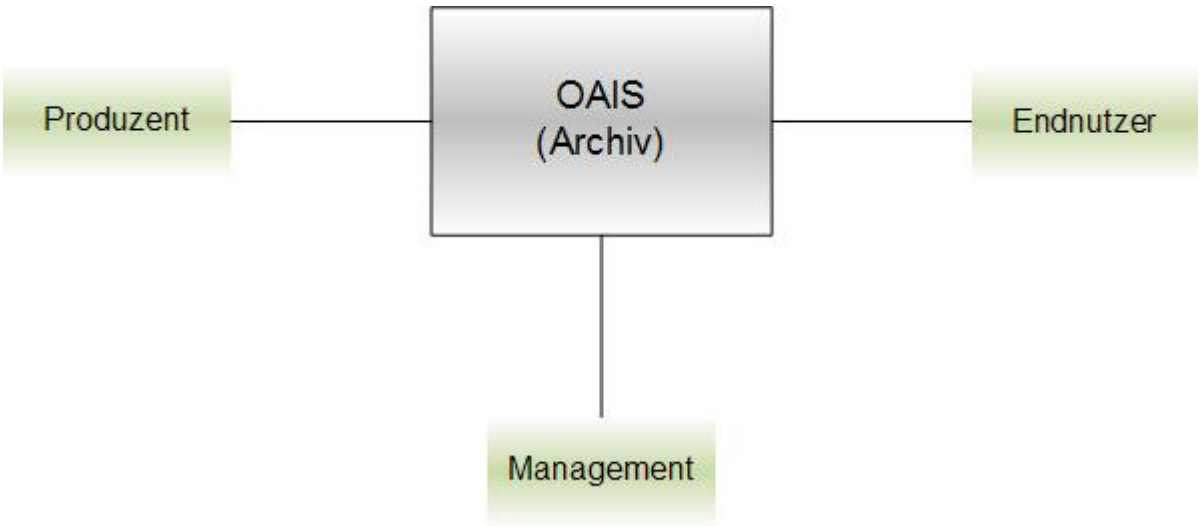

Abbildung 2-1: Umgebungsmodell eines OAIS

Außerhalb des OAIS gibt es **Produzenten**, **Endnutzer** und das **Management**.

- Die Rolle der Produzenten spielen diejenigen Personen oder Client-Systeme, welche die zu erhaltenden Information liefern.

- Die Rolle des Managements spielen diejenigen, welche die gesamte OAIS-Policy als Bestandteil einer umfassenderen Policy festlegen, zum Beispiel als Teil einer größeren Organisation. Mit anderen Worten ist die Management-Kontrolle des OAIS nur eine der Aufgaben des Managements. Das Management ist aber nicht mit den täglichen Betriebsabläufen des Archivs befasst. Die Aufgabe, das Tagesgeschäft zu führen, wird in der Funktionseinheit ‚Administration‘ in 4.1 näher beschrieben.

[10] Die deutsche Übersetzung folgt hier der englischen Vorlage, die allerdings abweichend von der OAIS-Definition im Glossar auf Abschnitt 6 verweist. Im Glossar wird noch auf die Erfüllung der in Abschnitt 4 definierten Verantwortlichkeiten verwiesen.

- Die Rolle der Endnutzer spielen diejenigen Personen oder Client-Systeme, die mit OAIS-Diensten Verbindung aufnehmen, um archivierte Information von Interesse aufzufinden und zu erlangen. Eine besondere Kategorie von Endnutzern ist die vorgesehene Zielgruppe. Die vorgesehene Zielgruppe ist derjenige Kreis von Endnutzern, die in der Lage sein sollte, die archivierte Information zu verstehen. Konkrete Personen oder Systeme können sowohl die Rolle des Endnutzers als auch die des Produzenten einnehmen.

Andere OAIS-Archive werden nicht explizit erläutert. Solche Archive können spezielle Vereinbarungen untereinander treffen, die mit den Anforderungen des Managements und des OAIS übereinstimmen. Andere Archive können mit einem bestimmten Archiv aus verschiedenen Gründen und gemäß unterschiedlich formalisierten, im Voraus festgelegten Vereinbarungen interagieren. Ein OAIS kann gegenüber einem anderen OAIS die Rolle des Produzenten übernehmen; zum Beispiel wenn die Verantwortung für die Archivierung bestimmter Informationstypen an dieses andere Archiv abgegeben werden muss. Ein OAIS kann aber auch die Rolle des Endnutzers eines anderen OAIS übernehmen; zum Beispiel wenn das erste OAIS sich entscheidet, sich für bestimmte Arten von selten gebrauchten Information auf das andere OAIS zu verlassen und sie nicht selbst zu archivieren. Eine solche Annahme sollte auf einer formalen Basis gründen, die auch festhält, dass jede Art von Policywechsel, die diese Annahme beeinträchtigen könnte, zwischen den Archiven mitgeteilt wird. Die verschiedenen Möglichkeiten zur Zusammenarbeit zwischen OAIS-Archiven werden in Abschnitt 6 „Archivische Interoperabilität“ behandelt.

2.2. OAIS Information

2.2.1. Definition von Information

Eine klare Definition von Information ist zentral für die Fähigkeit eines OAIS, diese Information zu erhalten. Während eine formale Informationsmodellierung in Abschnitt 4 vorgenommen wird, werden in diesem Unterabschnitt einige zentrale Konzepte beschrieben.

Eine Person oder ein System hat in der Regel ein **Grundwissen**, das es dieser Person oder diesem System ermöglicht, empfangene Information zu verstehen. Zum Beispiel wird eine Person, deren Grundwissen das Verständnis der englischen Sprache umfasst, in der Lage sein, einen englischen Text zu lesen und zu verstehen.

Information wird definiert als jede Art von austauschbarem Wissen, und diese Information wird beim Austausch immer durch eine Art von Daten ausgedrückt (d.h. repräsentiert). Zum Beispiel: Die Information in einem gedruckten Buch wird üblicherweise ausgedrückt durch sichtbare Buchstaben (die Daten), welche, verbunden mit der Kenntnis der verwendeten Sprache (dem Grundwissen), in eine bedeutungsvollere Information umgewandelt werden. Wenn das Grundwissen des Empfängers keine Englischkenntnisse umfasst, muss der englische Text (die Daten) von einem englischen Wörterbuch und einer Grammatik (d.h. von **Repräsentationsinformationen)** in einer Form, die für den Empfänger auf der Basis seines Grundwissens verstehbar sind, begleitet werden. Die vorgesehene Zielgruppe, für die die Information im Archiv erhalten wird, und ihr Grundwissen werden vom Archiv definiert. Dieses Grundwissen wird sich, wie weiter unten beschrieben, im Lauf der Zeit verändern. Die Definition der vorgesehenen Zielgruppe kann Gegenstand einer Vereinbarung zwischen den Geldgebern und anderen Anspruchsgruppen sein.

In ähnlicher Weise wird die in einer CD-ROM-Datei gespeicherte Information durch die Bits (die Daten) ausgedrückt, die, wenn sie mit der Repräsentationsinformation für diese Bits kombiniert werden, in eine verständlichere Information umgewandelt werden, so lange diese Repräsentationsinformation mit Hilfe des Grundwissens des Empfängers verstehbar ist. Nehmen wir zum Beispiel an, die Bits repräsentieren eine ASCII-Tabelle mit Zahlen, welche die in Breiten- und Längengraden gemessenen Koordinaten eines Orts auf der Erdoberfläche angeben. Die Repräsentationsinformationen schließen in diesem Fall typischerweise

ebenso eine Definition von ASCII ein wie eine Beschreibung des Formats der Zahlen und ihrer Verortung in der Datei, ihre Definitionen als Längen- und Breitengrade und ihre Definition als Gradeinheiten. Auch kann sie noch weitere, der Tabelle zugeordnete Bedeutungen enthalten. Ein anderes Beispiel für die Repräsentationsinformation einer eine FITS-Datei darstellenden Bitfolge könnte aus dem das Format definierenden FITS-Standard und einem Wörterbuch bestehen, in welchem die Bedeutung jener Schlüsselworte des Texts erklärt werden, die nicht Teil des Standards sind. Allgemein lässt sich sagen: „Daten, die unter Zuhilfenahme ihrer Repräsentationsinformation interpretiert werden, ergeben Information." Dies wird in Abbildung 2-2 schematisch dargestellt.

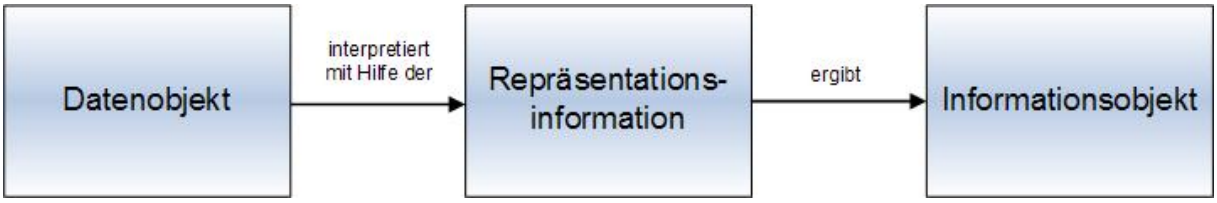

Abbildung 2-2: Informationsgewinnung aus Daten

Damit dieses **Informationsobjekt** erfolgreich archiviert werden kann, ist es von entscheidender Bedeutung für ein OAIS, das **Datenobjekt** und seine zugehörige Repräsentationsinformation klar zu identifizieren und eindeutig zu verstehen. Für digitale Information bedeutet das, dass das Archiv die Bits und die auf diese bezogenen Repräsentationsinformation genau bestimmen muss. Die erforderliche Transparenz auf der Bit-Ebene ist für die Erhaltung digitaler Information ein entscheidendes Element, sie läuft objektorientierten Konzepten zuwider, die diese Umsetzungsfragen zu verbergen versuchen. Für die Erhaltung digitaler Information bedeutet dies eine große Herausforderung.

Eine weitere Schwierigkeit ergibt sich aus der rekursiven Natur der Repräsentationsinformation, die typischerweise aus ihren eigenen Daten und anderer Repräsentationsinformation besteht, was typischerweise zu einem Netzwerk von Repräsentationsinformationsobjekten führt. Da es das Hauptziel eines OAIS ist, Information für eine vorgesehene Zielgruppe zu erhalten, muss das Archiv das Grundwissen seiner vorgesehenen Zielgruppe kennen, um zu wissen, welches Minimum an Repräsentationsinformationen gepflegt werden muss. Das OAIS muss sich also entscheiden, ob es eine minimale Repräsentationsinformation für seine vorgesehene Zielgruppe oder ob es ausführlichere Repräsentationsinformationen aufbewahren will, die von einem größeren Endnutzerkreis mit weniger speziellen Kenntnissen verstanden werden können, was einer Ausweitung der Definition der vorgesehenen Zielgruppe gleich kommen würde. Im Laufe der Zeit kann die Entwicklung des Grundwissens der vorgesehenen Zielgruppe es erfordern, die Repräsentationsinformationen zu aktualisieren, um die Verständlichkeit weiterhin sicherzustellen.

Die Wahl eines OAIS, alle relevante Repräsentationsinformation selbst zu sammeln oder auf ihr Vorhandensein in einem anderen vertrauenswürdigen oder Partner-OAIS-Archiv zu referenzieren, ist eine Implementierungs- und Organisations-Entscheidung.

In der Praxis verwendet man Software, die teilweise selbst Repräsentationsinformation sein kann, um Zugriff auf das Informationsobjekt zu bekommen, und diese Software beinhaltet auch ein gewisses Verständnis des Netzwerks der Repräsentationsinformationsobjekte. Jedoch sollte diese Software nicht als Begründung verwendet werden, um leicht verstehbare Repräsentationsinformationen, die das Informationsobjekt definieren, nicht zu bestimmen und zu sammeln. Denn es ist schwieriger, eine Software lauffähig zu halten als Information in digitaler oder in Papierform zu erhalten.

Das OAIS-Referenzmodell stellt den Erhalt des Informationsgehalts in den Vordergrund. Durch die sich weiter entwickelnde digitale Technologie werden die Multimediatechnologie und die Abhängigkeit vom komplexen Zusammenspiel zwischen Daten und Präsentationstechnologien einige Organisationen dazu bringen, zu verlangen, dass das Look and Feel der originalen Informationsdarstellung erhalten wird. Diese Art von Erhaltungsanforderungen kann es notwendig machen, Programme und Schnittstellen, die für den Zugriff auf die Daten gebraucht werden, zu erhalten. Das Problem kann durch die proprietäre Natur einiger Software weiter erschwert werden. Momentan werden verschiedene Techniken zur Erhaltung des Look and Feel beim Zugriff auf Information erforscht und mit Prototypen getestet. Diese Techniken, welche die Emulation auf der Hardware-Ebene, die Emulation von verschiedenen Programmierschnittstellen für allgemeine Dienste und die Entwicklung virtueller Maschinen mit einschließen, untersuchen die Erhaltung von originalem Bitstreams und Software über Technologien hinweg. Auch wenn sich das OAIS-Referenzmodell nicht auf diese aufkommenden Techniken konzentriert, dürfte es eine Architekturbasis für die prototypische Entwicklung dieser Techniken und deren Vergleich zur Verfügung stellen. Detaillierter werden die mit der Erhaltung des Look and Feel beim Informationszugriff zusammenhängenden Fragen im 5.2 angesprochen.

2.2.2 Definition eines Informationspakets

Die Definition eines Informationsobjekts ist auf alle Informationstypen anwendbar, über die in diesem und den nachfolgenden Unterabschnitten gesprochen wird. Mit anderen Worten haben alle Informationstypen zugehörige Repräsentationsinformationen, auch wenn diese normalerweise nicht explizit dargestellt werden.

Jede Übergabe von Information von einem Produzenten an ein OAIS und jede Auslieferung von Information an einen Endnutzer erscheint als eine mehr oder weniger eigenständige Übermittlung. Deshalb ist es sinnvoll, das Konzept eines **Informationspakets** zu definieren.

Ein Informationspaket ist ein konzeptioneller Container mit zwei Arten von Information: Der **Inhaltsinformation** und den **Erhaltungsmetadaten (PDI)**. Die Inhaltsinformation und die Erhaltungsmetadaten (PDI) sind durch die **Verpackungsinformation** miteinander verbunden und identifizierbar. Das entstandene Paket kann mit Hilfe von **Erschließungsinformationen** wieder aufgefunden werden.

In Abbildung 2-3 werden diese Beziehungen innerhalb des Informationspakets schematisch dargestellt.

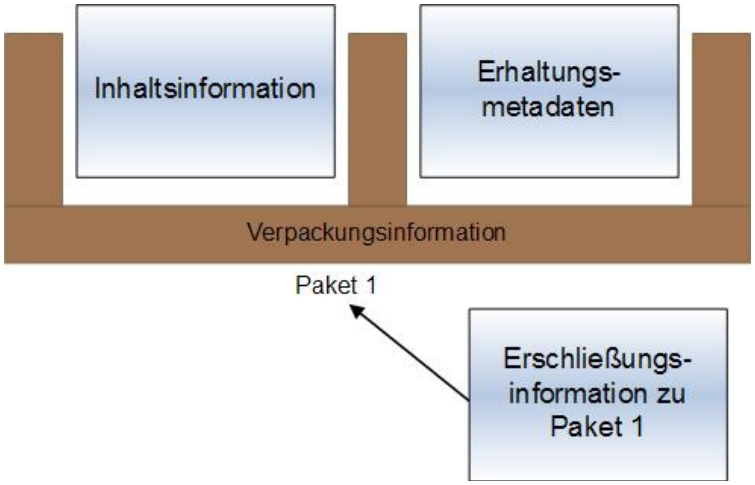

Abbildung 2-3: Konzepte und Beziehungen innerhalb eines Informationspakets

Die Inhaltsinformation ist das eigentliche Ziel der Erhaltung. Sie bestehen aus dem Inhaltsdatenobjekt (physisches Objekt oder digitales Objekt, d.h. Bits) und den zugehörigen Repräsentationsinformationen, die benötigt werden, um das Inhaltsdatenobjekt für die vorgese-

hene Zielgruppe verstehbar zu machen. Zum Beispiel kann das Inhaltsdatenobjekt ein Bild sein, das als Bit-Inhalt einer CD-ROM-Datei zusammen mit anderen, die Repräsentationsinformationen enthaltenden, Dateien auf ein und derselben CD-ROM liegt.

Nur wenn die Inhaltsinformation eindeutig definiert ist, können die Erhaltungsmetadaten festgelegt werden. Die Erhaltungsmetadaten beziehen sich auf die Inhaltsinformation. Sie werden gebraucht, um die Inhaltsinformation zu erhalten, um sicherzustellen, dass sie eindeutig identifiziert wird und um die Umgebung zu verstehen, in der die Inhaltsinformation erzeugt worden ist. Die Erhaltungsmetadaten sind in fünf Kategorien von Archivierungsinformationen eingeteilt: Provenienz, Kontext, Referenz, Beständigkeit und Zugriffsrechte:

- Provenienz beschreibt die Herkunft der Inhaltsinformation, wer sie seit ihrer Entstehung in Obhut hatte und ihre Geschichte (einschließlich der Bearbeitungsgeschichte).
- Kontext beschreibt, in welchem Verhältnis die Inhaltsinformation zu anderer Information außerhalb des Informationspakets steht. Zum Beispiel beschreiben die Kontextinformationen, warum die Inhaltsinformation erzeugt wurde. Außerdem können sie eine Beschreibung darüber enthalten, in welcher Beziehung sie zu einem anderen verfügbaren Inhaltsinformationsobjekt steht.
- Referenz enthält einen oder mehrere Identifikatoren oder ein System von Identifikatoren, durch welche die Inhaltsinformation eindeutig identifiziert werden kann. Beispielsweise die ISBN eines Buchs oder eine Reihe von Attributen, die eine Instanz von Inhaltsinformation von einem anderen unterscheiden.
- Beständigkeit bietet eine Art Hülle oder Schutzschild, welche(s) die Inhaltsinformation vor undokumentierter Veränderung schützt. Das kann zum Beispiel eine Prüfsumme über die Inhaltsinformation eines digitalen Informationspakets sein.
- Zugriffsrechte enthalten die Zugriffsbedingungen einschließlich Erhaltung, Verbreitung und Gebrauch der Inhaltsinformation. Zum Beispiel würden sie die Bewilligungen enthalten, welche dem OAIS Erhaltungsmaßnahmen erlauben, Lizenzierungsangebote (zum Vertrieb) und Spezifikationen sowohl für Maßnahmen zur Durchsetzung der (Nutzungs-)Rechte wie auch zur Zugriffskontrolle.

Die Verpackungsinformation ist diejenige Information, die, entweder tatsächlich oder logisch, die Inhaltsinformation und die Erhaltungsmetadaten verknüpft, identifiziert und zueinander in Beziehung setzt. Wenn zum Beispiel die Inhaltsinformation und die Erhaltungsmetadaten als Inhalt konkreter Dateien einer CD-ROM identifiziert sind, dann würde die Verpackungsinformation sowohl die ISO 9660 Struktur von Datenträger und Dateien der CD-ROM als auch die Namen und die Verzeichnisinformationen der Dateien auf der CD-ROM beinhalten. Ein anderes Beispiel zur Paketierung ist XFDU (s. Referenz D 12[11]), wo die Verpackungsinformation aus dem Identifikator der Datei und der Definition der Paketierungsmethode besteht.

Die Erschließungsinformation ist die Information, die verwendet wird, um herauszufinden, welches Paket die gesuchte Inhaltsinformation enthält. Je nach Umgebung kann es sich dabei entweder nur um einen aussagekräftigen Titel des Informationspakets handeln, der in einer Nachricht auftaucht, oder es kann sich um einen ganzen Satz von Attributen handeln, der in einem Katalogdienst durchsucht werden kann.

[11] Die Anhänge sind in der deutschen Übersetzung nicht enthalten.

2.2.3. Varianten von Informationspaketen

Es ist notwendig, zwischen einem Informationspaket, das von einem OAIS aufbewahrt wird, und Informationspaketen, die an ein OAIS übergeben oder von dort wieder ausgeliefert werden, zu unterscheiden. Diese Paketvarianten werden gebraucht, um der Tatsache gerecht zu werden, dass manche Übergaben an ein OAIS hinsichtlich der Anforderungen einer OAIS-konformen Aufbewahrung nur unzureichende Repräsentationsinformationen oder Erhaltungsmetadaten haben werden. Außerdem kann es sein, dass diese Übergaben ganz anders zusammengestellt sind als dies das OAIS mit der Information, die es aufbewahrt, tut. Schließlich kann es auch vorkommen, dass das OAIS Inhaltsinformation an Endnutzer heraus gibt, die nicht alle damit verknüpften Repräsentationsinformationen oder alle Erhaltungsmetadaten enthalten. Diese Varianten werden Übergabeinformationspaket (SIP), Archivinformationspaket (AIP) und Auslieferungsinformationspaket (DIP) genannt.

Das **Übergabeinformationspaket (SIP)** ist das Paket, was von einem Produzenten an das OAIS geschickt wird. Seine Form und sein genauer Inhalt werden typischerweise zwischen dem Produzenten und dem Archiv ausgehandelt (siehe die entsprechenden Standards in 1.5). Die meisten SIPs werden einige Inhaltsinformationen und einige Erhaltungsmetadaten enthalten.

Die Beziehungen zwischen SIPs und AIPs können sich komplex gestalten: Genauso wie eine einfache Eins-zu-eins-Beziehung, in der ein SIP ein AIP hervorbringt, schließen andere Möglichkeiten ein, dass ein AIP aus einer Vielzahl von SIPs, die zu verschiedenen Zeiten von einem oder vielen Produzenten hergestellt wurden, gebildet wird, dass aus einem SIP einer Mehrzahl von AIPs gebildet werden, oder dass viele SIPs aus einer oder mehreren Quellen entpackt und auf unterschiedliche Weise zu vielen AIPs zusammengesetzt werden. Selbst im ersten Fall kann das OAIS eine bestimmte Zahl an Transformationen am SIP vornehmen müssen. Die Verpackungsinformation wird dabei stets in einer gewissen Form präsent sein.

Innerhalb des OAIS werden ein oder mehrere SIPs für die Erhaltung in ein oder mehrere **Archivinformationspakete (AIPs)** verwandelt. Das AIP enthält einen kompletten Satz an Erhaltungsmetadaten für die zugehörige Inhaltsinformation. Das AIP kann auch eine Kollektion anderer AIPs enthalten, was in Abschnitt 4 diskutiert und modelliert wird. Die Verpackungsinformation des AIPs entspricht den jeweiligen OAIS internen Standards und sie kann sich verändern, während sie durch das OAIS verwaltet wird.

Als Antwort auf eine Anfrage stellt das OAIS ein ganzes oder Teile eines AIPs dem Endnutzer in Form eines **Auslieferungsinformationspakets (DIP)** zur Verfügung. Das DIP kann auch Kollektionen von AIPs enthalten und es kann alle oder nur einen Teil der Erhaltungsmetadaten umfassen. Die Verpackungsinformation ist notwendigerweise immer in irgendeiner Form vorhanden, damit der Endnutzer die von ihm abgefragte Information auch klar erkennen kann. Je nach Art der für die Auslieferung verwendeten Datenträger und je nach Bedürfnissen des Endnutzers kann die Verpackungsinformation verschiedene Formen annehmen.

2.3 Außenbeziehungen des OAIS im Überblick

Die folgenden Unterabschnitte präsentieren einen Überblick über das Zusammenspiel der einzelnen Einheiten in der OAIS-Umgebung. Abbildung 2-4 ist ein Datenflussdiagramm, das die operativen externen Datenflüsse des OAIS-Archivs darstellt. Dieses Diagramm konzentriert sich auf den Informationsfluss zwischen Produzenten, Endnutzern und dem OAIS und berücksichtigt nicht die Datenflüsse, welche das Management betreffen. Diese werden in Abschnitt 4 näher behandelt.

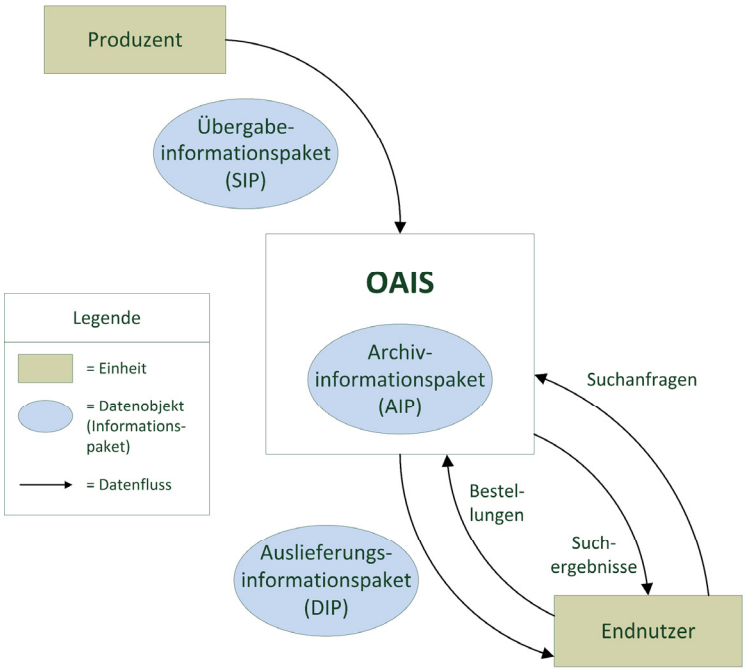

Abbildung 2-4: Externe Daten des OAIS-Archivs

2.3.1. Interaktion mit dem Management

Das Management gibt dem OAIS eine Satzung und einen Zuständigkeitsbereich. Die Satzung kann vom Archiv entwickelt werden, es ist aber wichtig, dass das Management die Archivaktivitäten formell bestätigt. Der Zuständigkeitsbereich bestimmt die Produzenten und Endnutzergruppen, die vom Archiv bedient werden.

Einige Beispiele für typische Interaktionen zwischen OAIS und Management:

- Das Management ist häufig die Hauptfinanzierungsquelle eines OAIS und kann Richtlinien für die Verwendung der Mittel erlassen (Personal, Material, Einrichtung).
- Das Management wird im Allgemeinen regelmäßig die Leistung des OAIS und seine Fortschritte hinsichtlich langfristiger Ziele evaluieren und die Risiken, denen das OAIS und seine Bestände ausgesetzt sind, einschätzen.
- Das Management bestimmt gegebenenfalls die Preispolitik für Dienste des OAIS oder es bestätigt sie zumindest.
- Das Management beteiligt sich an der Lösung von Konflikten zwischen Produzenten, Endnutzern und der internen Verwaltung des OAIS.

Ein effektives Management sollte das OAIS auch dadurch unterstützen, dass es Abläufe etabliert, die sicherstellen, dass das OAIS innerhalb seines Einflussbereichs auch in Anspruch genommen wird. Zum Beispiel sollte die Strategie des Managements vorsehen, dass alle Einheiten, die von ihm finanziert werden und sich innerhalb seines Einflussbereichs befinden, Datenprodukte an das Archiv übermitteln und sich an die Archivstandards und –abläufe halten.

2.3.2 Interaktion mit den Produzenten

Der erste Kontakt zwischen OAIS und Produzent ist der Antrag, dass das OAIS die vom Produzenten erzeugten Datenprodukte aufbewahrt. Dieser Kontakt kann vom OAIS, vom Produzenten oder vom Management ausgehen. Der Produzent trifft eine **Übergabevereinbarung** mit dem OAIS, welche die zu übergebenden SIPs eindeutig bestimmt und einen beliebigen Zeitrahmen für diese Übergabe abstecken kann. Einige Über-

gabevereinbarungen werden die zwingende Verpflichtung, dem OIAS Information zu übergeben, widerspiegeln, während andere eine freiwillige Anbietung der Information widerspiegeln werden, und wieder andere werden jegliche Zahlungen, die vorkommen können, wiedergeben. Auch wenn keine formelle Übergabevereinbarung existiert, wie z.B. bei der Archivierung zahlreicher Seiten aus dem World Wide Web, könnte eine virtuelle Übergabevereinbarung existieren, die die Dateiformate spezifiziert und die Inhalte, die das OAIS akzeptiert.

Innerhalb der Übergabevereinbarungen werden eine oder mehrere **Datenübergabesitzungen** bestimmt. Es können beträchtliche Zeitabstände zwischen den Datenübergabesitzungen liegen. Eine Datenübergabesitzung kann ein oder mehrere SIPs enthalten, die als ein Satz Datenträger oder in einer einzelnen Onlineübertragung geliefert werden. Der Inhalt einer Datenübergabesitzung basiert auf einem Datenmodell, das zwischen Produzent und OAIS in der Übergabevereinbarung ausgehandelt wurde. Dieses Datenmodell bestimmt die logischen Komponenten eines SIPs (z.B. die Inhaltsdatenobjekte, Repräsentationsinformation, Erhaltungsmetadaten, Verpackungsinformation und Erschließungsinformation), die abgeliefert werden müssen, und wie (und ob) sie in jeder Datenübergabesitzung enthalten sind. Alle Datenlieferungen im Rahmen einer Übergabevereinbarung werden als zu dieser Übergabevereinbarung gehörig erkannt und besitzen in der Regel ein konsistentes Datenmodell, das in der Übergabevereinbarung beschrieben wird. Zum Beispiel kann eine Datenübergabesitzung aus einem Satz an Inhaltsinformationen bestehen, die einem Satz von Beobachtungen entsprechen und die in Form einer Anzahl von Dateien auf einer CD-ROM transportiert werden. Die Erhaltungsmetadaten sind in diesem Beispiel zwischen zwei andere Dateien aufgeteilt. Alle diese Dateien brauchen Repräsentationsinformationen, die auf irgendeine Weise bereitgestellt werden müssen. Die CD-ROM und deren Datei- und Verzeichnisstruktur sind die Verpackungsinformation, die die Einkapselung und Identifizierung der Inhaltsinformation und der Erhaltungsmetadaten bei der Datenübergabesitzung ermöglichen. Die Übergabevereinbarung regelt, wie die Repräsentationsinformation für jede einzelne Datei bereitgestellt wird, wie die CD-ROM erkannt werden kann, wie die Verpackungsinformation genutzt wird, um die Inhaltsinformation und die PDI des SIPs zu bestimmen und zusammenzuhalten sind und wie häufig Datenübergabesitzungen stattfinden werden (z.B. zwei Jahre lang einmal monatlich). Außerdem enthält die Vereinbarung weitere benötigte Information wie z.B. Zugriffsbeschränkungen auf die Daten und Vorgaben zu deren Durchsetzung.

Jedes SIP einer Datenübertragungssitzung muss Mindestanforderungen des OAIS hinsichtlich Vollständigkeit erfüllen. Trotzdem kann es in einzelnen Fällen notwendig sein, mehrere SIPs entgegen zu nehmen, bevor ein brauchbares Archivinformationspaket erzeugt und ins OAIS übernommen werden kann. In anderen Fällen kann ein einzelnes SIP Daten enthalten, die zahlreichen AIPs beigegeben werden müssen. Eine Übergabevereinbarung beinhaltet (oder referenziert) außerdem die Verfahren und Protokolle, anhand derer ein OAIS entweder den Erhalt und die Vollständigkeit einer Datenübergabesitzung mit dem Produzenten verifiziert oder sich beim Produzenten nach den Inhalten der Datenübergabesitzung erkundigt.

2.3.3 Interaktion mit den Endnutzern

Es gibt viele verschiedene Formen der Interaktion zwischen den Endnutzern und dem OAIS. Sie umfassen Fragen an einen Helpdesk, Literaturanfragen, Katalogrecherchen, Bestellungen oder Anfragen nach dem Status einer Bestellung. Der Bestellprozess ist für das OAIS-Referenzmodell von besonderem Interesse, weil er sich mit dem Austausch von Archivbeständen zwischen dem OAIS und dem Endnutzer beschäftigt.

Der Endnutzer schließt eine **Bestellvereinbarung** für Information mit dem OAIS ab. Diese Information kann bereits im Archiv vorhanden sein oder wird für die künftige Übernahme erwartet. Die Bestellvereinbarung kann jede beliebige Zeitspanne umfassen und während ihrer Gültigkeit können eine oder mehrere **Datenauslieferungssitzungen** stattfinden. Eine Datenauslieferungssitzung kann den Transfer eines Satzes von Datenträgern oder eine ein-

zelne Onlineübertragung beinhalten. Die Bestellvereinbarung identifiziert ein oder mehrere gewünschte AIPs, legt fest, wie diese AIPs umgewandelt und in Auslieferungsinformationspakete (DIPs) abgebildet werden und wie diese DIPs in einer Datenauslieferungssitzung verpackt werden. In der Bestellvereinbarung werden außerdem andere wichtige Information angegeben wie Lieferangaben (z.B. Namen oder Mailadresse), rechtliche Angaben (z.B. Nutzungsbeschränkungen, autorisierte Endnutzer oder Lizenzgebühren) sowie ggf. die Preise. Es gibt zwei übliche Bestellarten durch Endnutzer: Die **ereignisbasierte Bestellung** und die **Spontan-Bestellung**.

Im Fall einer Spontan-Bestellung schließt der Endnutzer für im Archiv verfügbare Information eine Bestellvereinbarung mit dem OAIS ab. Wenn der Endnutzer nicht von vornherein weiß, welche der vorhandenen Bestände des OAIS für ihn von Interesse sind, wird der Endnutzer zunächst eine **Recherchesitzung** im OAIS starten. Während dieser Recherchesitzung nimmt der Endnutzer die **Findmittel** des OAIS zu Hilfe, die auf den **Erschließungsinformationen** oder in einigen Fällen auf den AIPs selber basieren, um mögliche Bestände, die für ihn von Interesse sind, zu identifizieren und zu recherchieren. Das kann durch die Übermittlung von Suchanfragen und die Rückgabe von Suchanfragen (z.B. Ergebnislisten) an den Endnutzer erreicht werden. Dieser Rechercheprozess neigt dazu, iterativ zu sein, wenn der Endnutzer zunächst grobe Suchkriterien verwendet und dann aufgrund erster Suchergebnisse diese Kriterien verfeinert. Sobald der Endnutzer dem OAIS die AIPs seines Interesses nennt, kann er eine Bestellvereinbarung vorlegen, welche die Identifikatoren der gewünschten AIPs enthält sowie Angaben darüber, wie er die DIPs vom OAIS übernehmen will. Wenn die AIPs verfügbar sind, wird eine Spontan-Bestellung durchgeführt. Wenn aber die gewünschten AIPs noch nicht verfügbar sind, kann eine ereignisbasierte Bestellung erfolgen.

Im Fall einer ereignisbasierten Bestellung schließt der Endnutzer eine Bestellvereinbarung mit dem OAIS ab für Information, die er beim Eintreffen eines bestimmten Ereignisses zu erhalten erwartet. Dieses Ereignis kann regelmäßig stattfinden, z.B. eine monatliche Übermittlung aller AIPs, welche das OAIS von einem bestimmten Produzenten übernommen hat. Es kann sich auch um ein einmaliges Ereignis handeln, wie z.B. die Übernahme oder die Erstellung eines spezifischen AIPs. Die Bestellvereinbarung wird außerdem andere notwendige Information festlegen, so z.B. das auslösende Ereignis für neue Datenauslieferungssitzungen und die Kriterien für die Auswahl der OAIS-Bestände, die in jede neue Datenauslieferungssitzung aufzunehmen sind.

Die Bestellvereinbarung muss kein formelles Dokument sein. In der Regel wird ein OAIS eine allgemeine Preispolicy und eine Datenbank mit den elektronischen und physischen Kontaktdaten seiner Benutzer haben. In diesem Fall kann das Verfassen einer Bestellvereinbarung darin bestehen, ein Internetformular mit den Angaben eines interessierenden AIPs auszufüllen.

3 OAIS-Aufgabenbereiche

Unterabschnitt 3.1. nennt die Minimalanforderungen, die ein OAIS erfüllen sollte. Unterabschnitt 3.2. gibt einige Beispiele von Mechanismen, um diese Aufgaben zu erfüllen, obwohl nicht alle davon in allen OAIS-Archiven anwendbar sein werden, und 1.5 bietet Hinweise zu einigen relevanten Standards.

3.1 Verbindliche Aufgaben

Dieser Unterabschnitt legt die verbindlichen Aufgaben dar, die eine Organisation erfüllen muss, um ein OAIS zu betreiben.

Das OAIS sollte:

- mit Informations-Produzenten über Information verhandeln und diese entsprechend annehmen.

- genügend Kontrolle über die angebotene Information bekommen, in dem Maß, das benötigt wird, um deren Langzeiterhaltung sicherzustellen.

- bestimmen, entweder alleine oder zusammen mit anderen, welche Gruppen zur vorgesehenen Zielgruppe gehören sollten und deswegen fähig sein sollten die angebotene Information zu verstehen, um dadurch ihr Grundwissen zu definieren.

- sicherstellen, dass die zu erhaltende Information für die vorgesehene Zielgruppe **unmittelbar verstehbar** ist. Insbesondere sollte die vorgesehene Zielgruppe befähigt sein, die Information ohne den Gebrauch spezieller Hilfsmittel wie die Hilfe von Experten, die die Information erstellt haben, zu verstehen.

- dokumentierten Richtlinien und Abläufen folgen, die sicherstellen, dass die Information gegen alle vorstellbaren Gefahren geschützt sind, einschließlich der Schließung eines Archivs, sicherstellend, dass sie niemals gelöscht wird, außer wenn es als Bestandteil einer erprobten Strategie gestattet wird. Es sollte keine ad-hoc Löschungen geben.

- die archivierten Information der vorgesehenen Zielgruppe verfügbar machen und die Auslieferung der Information ermöglichen, als Kopien der ursprünglich übergebenen Datenobjekten, oder zu diesen zurückverfolgbar, mit Belegen für ihre Authentizität.

3.2 Beispielhafte Mechanismen, wie man die Aufgaben erfüllen kann

Dieser Unterabschnitt enthält beispielhafte Mechanismen, um die in 3.1 formulierten Aufgaben zu erfüllen. Nicht alle dieser Mechanismen werden von allen OAIS anwendbar sein.

3.2.1 Verhandeln und Annehmen von Information

Eine Organisation, die ein OAIS betreibt, sollte einige Kriterien eingeführt haben, die dabei helfen, zu bestimmen, welche Arten der Information das Archiv annehmen möchte oder muss. Diese Kriterien können unter anderem einschließen: das Thema der Information, deren Quelle, den Grad der Einmaligkeit oder Originalität und die Art der Techniken, mit denen die Information abgebildet wird (physischer Datenträger, digitaler Datenträger, Format). Die Information kann überhaupt in einer großen Vielfalt von gewöhnlichen und nicht so gewöhnlichen Formen, wie Büchern, Dokumenten, Karten, Datensätzen und physischen Objekten und unter Benutzung vieler Kommunikationswege wie Netzwerke, Mails und Sonderzustellung übergeben werden.

Das OAIS sollte mit dem Produzenten verhandeln, um sicherzustellen, dass es angemessene Inhaltsinformationen erhält und die dazugehörigen Erhaltungsmetadaten, die seinem Auftrag und der vorgesehenen Zielgruppe entsprechen. Beträchtliche Iterationen können notwendig sein, um sich zu einigen, welches die richtige Information ist, die übergeben werden, und wie sie in die Form kommen, die für ein OAIS akzeptabel ist. Zum Beispiel kann es sein, dass Audio- oder Videomaterial digitalisiert wird oder dass ein Text gescannt wird.

Verhandlungen können eine Palette an möglichen Aktionen umfassen, und manchmal zu keinem Ergebnis führen. Sie könnten für jedes SIP, jeden Produzenten oder jede Gruppe von Produzenten durchgeführt werden. Sie beinhalten vielleicht extensive menschliche Kommunikation oder Abstimmungen zwischen Systemen, etwa solchen wie sie zwischen einem Webcrawler und einem Webserver stattfinden, wenn zu entscheiden ist, welche Inhalte, falls überhaupt, der Server dem Crawler erlaubt zu übernehmen (Ersterer agiert als Produzent, Letzterer als OAIS).

Das OAIS sollte genügend Erschließungsinformationen extrahieren oder anderweitig bekommen, um der vorgesehenen Zielgruppe zu helfen, die für sie interessante Inhaltsinformation zu finden. Es sollte auch sicherstellen, dass die Information alle OAIS-internen Standards erfüllen.

3.2.2 Ausreichende Kontrolle für die Langzeitarchivierung erlangen

Es ist wichtig, dass das OAIS den Unterschied erkennt, der zwischen dem physischem Eigentum oder Besitz der Inhaltsinformation und dem Besitz des geistigen Eigentums[12] an dieser Information bestehen kann. Wenn es diese Information erzeugt hat und der gesetzliche Eigentümer der Inhaltsinformation ist, dann hat das OAIS die Freiheit, alles zu tun, was nötig ist, um die Information zu erhalten und zugänglich zu machen. Wenn es die Inhaltsinformation von irgendeinem anderen Produzenten oder einer Institution bekommt, dann sollte das OAIS sicherstellen, dass es eine gesetzlich gültige Transfervereinbarung gibt, die entweder das geistige Eigentum an das OAIS überträgt oder ganz klar die Rechte definiert, die dem OAIS zugestanden werden und Beschränkungen, die der bzw. die Rechteinhaber ihm auferlegen. Das OAIS sollte sicherstellen, dass die nachfolgenden Aktionen zur Erhaltung und Verfügbarkeit diesen Rechten und Beschränkungen entspricht. Wenn das OAIS das Recht am geistigen Eigentum nicht bekommt, dann sollte der Vertrag genau definieren, inwiefern der/die Rechteinhaber bei der Erhaltung, dem Management oder der Herausgabe der Information beteiligt sind. In den meisten Fällen wird es für das OAIS besser sein, einen Vertrag auszuhandeln, der die Forderungen des/der Rechteinhaber/s genau definiert und das OAIS autorisiert, in Übereinstimmung mit diesen Bestimmungen zu handeln, ohne aktive Beteiligung des/der Rechteinhaber/s in speziellen Fällen.

Das OAIS muss eine ausreichende Kontrolle über die Inhaltsinformation und die Erhaltungsmetadaten übernehmen, damit es sie langfristig erhalten kann. Das betrifft nicht die Verpackungsinformation der AIPs, weil diese per Definition sowieso innerhalb des OAIS kontrolliert wird. Die Problemfelder, wenn man ausreichende Kontrolle über vorwiegend digitale Inhaltsinformation und Erhaltungsmetadaten übernehmen will, werden in drei verwandten Kategorien, wie folgt, behandelt:

- Urheberrecht[13], geistiges Eigentum und andere gesetzliche Beschränkungen hinsichtlich der Benutzung;
- Das Recht, die Repräsentationsinformation zu ändern;
- Vereinbarungen mit externen Organisationen

[12] Der Begriff der „Intellectual Property Rights“ lässt sich mit „geistigem Eigentum“ eigentlich nur umschreiben, da das anglo-amerikanische Urheberrecht auf anderen Konzepten beruht als das deutsche. Im Grunde geht es im Folgenden um die Verwertungsrechte des Urhebers.
[13] Hier wird im Original auf das “Copyright” abgehoben.

Urheberrecht, geistiges Eigentum und andere gesetzliche Beschränkungen: Ein Archiv wird alle maßgeblichen gesetzlichen Beschränkungen berücksichtigen. Diese Fragen tauchen auf, wenn das OAIS als Treuhänder handelt. Ein OAIS sollte das Konzept des Rechts am geistigen Eigentum, wie das Urheberrecht, und alle maßgeblichen anderen Gesetze verstehen, bevor es urheberrechtlich geschütztes Material ins OAIS aufnimmt. Falls nötig, kann es Richtlinien für die Übernahme von Information und Regeln für die Auslieferung und Duplizierung der Information aufstellen. Außerhalb des Geltungsbereichs dieses Dokuments sind Details des nationalen und Internationalen Urheberrechts zu berücksichtigen.

Das Recht, Inhaltsinformationen zu ändern: Obwohl die Persistenzinformation in den Erhaltungsmetadaten eines AIPs sicherstellen, dass die Inhaltsinformation-bezogenen Bits nicht geändert werden sollen, wird eine Zeit kommen, in der die Bits der Inhaltsinformation in einer Form sind, die für die Endnutzer der vorgesehenen Zielgruppe ungeeignet sind. Die Bits der Inhaltsinformation können vollständig in einem Papierausdruck dokumentiert sein, sodass die Information technisch nicht verloren gegangen ist, aber praktisch wird sie nicht verfügbar sein. Das OAIS benötigt die Befugnis, die Inhaltsinformation in eine neue Repräsentationsform zu übertragen. Wenn es als Treuhänder agiert, kann es notwendig sein, weitere Befugnisse einzuholen, um solche Änderungen vorzunehmen. Wenn die Information urheberrechtlich geschützt ist, sollte das OAIS bereits im Vorfeld eine Erlaubnis für die Änderungen, die nötig sind, um Erhaltungsziele zu erfüllen, ausgehandelt haben. Es könnte OAIS-externe Fachexperten beschäftigen, um zu gewährleisten, dass Informationen nicht verloren gehen. Im Idealfall werden in solch einer Situation die originalen AIPs (vollständig beschrieben) und die neuen AIPs aufbewahrt werden. Fragen der digitalen Migration werden in 5.1 näher besprochen.

Vereinbarungen mit externen Organisationen: Ein OAIS kann mit anderen Organisationen eine Reihe Vereinbarungen treffen, um seine Erhaltungsziele zu unterstützen. Zum Beispiel kann es mit einem anderen OAIS eine Kooperation schließen, damit es nicht alle geläufigen Repräsentationsinformationsobjekte zu ihren Inhaltsinformationsobjekten aufbewahren muss. Die Einhaltung und Zweckmäßigkeit solcher Vereinbarungen mit anderen Organisationen sollte beobachtet werden.

3.2.3 Vorgesehene Zielgruppe bestimmen

Die Übergabe, oder geplante Übergabe, von Inhaltsinformationen und verbundenen Erhaltungsmetadaten erfordert die Bestimmung der erwarteten Endnutzer, oder der vorgesehenen Zielgruppe, dieser Information. Dies ist notwendig, um zu bestimmen, ob die Information, so wie sie ist, für diese Zielgruppe verstehbar sein wird. Ein Archiv kann beispielsweise entschieden, dass bestimmte Inhaltsinformationen für die allgemeine Öffentlichkeit verstehbar sein sollen, sodass diese die vorgesehene Zielgruppe wird.

Für manche Wissenschaftsdaten könnte die vorgesehene Zielgruppe beschrieben sein mit solchen Endnutzern, die den Ausbildungsstand des ersten Jahrs eines Hochschulstudiums in einer verwandten wissenschaftlichen Disziplin erreicht haben. Dies ist ein schwierigerer Fall, da weniger klar ist, welches Maß an spezieller Wissenschaftsterminologie tatsächlich angemessen sein kann. Die Produzenten solcher speziellen Information sind oftmals mit einer begrenzt bekannten Terminologie vertraut, sodass es besonders entscheidend ist, die vorgesehene Zielgruppe für die Information klar zu definieren und sich zu bemühen, dass diese Zielgruppe die Information in jedem Fall versteht.

Mögliche Veränderungen der vorgesehenen Zielgruppe sind ebenfalls zu berücksichtigen. Information, die ursprünglich für eine eng begrenzte Zielgruppe vorgesehen war, kann irgendwann einmal für eine größere Gruppe verstehbar gemacht werden müssen. Zum Beispiel könnte Information, die ursprünglich für eine bestimmte wissenschaftliche Gemeinschaft verstehbar war, auch der allgemeinen Öffentlichkeit verstehbar gemacht werden müssen. Das bedeutet wahrscheinlich, dass den Repräsentationsinformationen und den Erhaltungsmetadaten Erklärungen zugefügt werden, und es kann zunehmend schwierig werden,

diese Information zu bekommen. Wird die vorgesehene Zielgruppe schon zu Beginn der Langzeiterhaltung breiter definiert, z.B. als allgemeine Öffentlichkeit, kann das Problem reduziert und die Wahrscheinlichkeit verbessert werden, dass die Information für alle in der ursprünglichen Nutzergruppe verstehbar sein wird.

3.2.4 Sicherstellen der unmittelbaren Verstehbarkeit von Information

Das Ausmaß, in welchem die Inhaltsinformation und die dazugehörigen Erhaltungsmetadaten Informationen an die vorgesehene Zielgruppe übermitteln, ist im Allgemeinen ziemlich subjektiv. Nichtsdestotrotz ist es unabdingbar, dass ein Archiv diese Bestimmung vornimmt, um die Informationserhaltung zu optimieren. Digitale Inhaltsinformationen und Erhaltungsmetadaten benötigen adäquate Repräsentationsinformationen, um für die vorgesehene Zielgruppe unmittelbar verstehbar zu sein. Typischerweise sind vielfältige Repräsentationsinformationsobjekte involviert, und das wird in 4.2 diskutiert.

Ein Beispiel: Stellen Sie sich eine Inhaltsinformation aus der digitalen Wetteraufzeichnung des Regen, der Temperatur, des Luftdrucks, und weiterer Parameter vor, die überall auf der Welt ein Jahr lang gemessen werden. Diese Art von Information ist sehr umfangreich und meist nicht in einer Form, die üblicherweise schnell durchgesehen oder gelesen werden kann, sondern in einer Form, die nur von einer speziellen Anwendungssoftware bearbeitet und gesucht werden kann. Diese Inhalte sind vielleicht nur für die ursprünglichen Produzenten selbst verstehbar, es sei denn, sie werden durch eine entsprechende Dokumentation über die Bedeutung der verschiedenen Wetteraufzeichnungen und ihrer Beziehung untereinander ergänzt sowie dadurch, wie die Werte zum Messinstrument zurück verfolgt werden. In solchen speziellen Gebieten braucht es eine besondere Anstrengung, um sicherzustellen, dass die Inhaltsinformation und die Erhaltungsmetadaten für die vorgesehene Zielgruppe verstehbar sind. Wenn das Archiv nicht über solches Spezialwissen verfügt, könnte es externe Vertreter der Nutzergruppe die Information auf langfristige Verstehbarkeit begutachten lassen. Sonst könnten diese Information möglicherweise nur von wenigen Spezialisten verstanden werden und würden verloren gehen, wenn diese Spezialisten nicht mehr zur Verfügung stehen.

Auch wenn ein Satz an Informationen als verstehbar für eine gewisse vorgesehene Zielgruppe bestimmt worden sind, kann es sein, dass das Grundwissen dieser Gruppe sich über die Jahre so entwickelt, dass wichtige Information nicht mehr leicht verstehbar ist. An diesem Punkt kann es für ein OAIS notwendig sein, die zugehörigen Repräsentationsinformationen zu ergänzen, damit sie für die vorgesehene Zielgruppe wieder allgemein verstehbar ist.

Ein anderes Beispiel: Eine Inhaltsinformation eines Manuskripts könnte auf Englisch geschrieben sein und daher sein Inhalt für ein großes Publikum allgemein verstehbar. Doch wenn der Grund, aus dem es geschrieben worden ist, nicht klar dokumentiert ist, kann viel von seiner Bedeutung verloren gehen. Dieser Entstehungszweck ist deswegen Teil seines Kontextes und muss in die Erhaltungsmetadaten aufgenommen werden.

Software wird für einen effizienten Zugriff auf digitale Inhaltsinformationen benötigt. Jedoch ist die Erhaltung von Software für spezifische Inhaltsinformationen über einen langfristigen Zeitraum aufgrund der begrenzten Anwendbarkeit solcher Software nicht kosteneffektiv. Die Gefahr des Datenverlusts ist groß, wenn man sich für die Erhaltung und Lesbarkeit von Information auf solche Software verlässt, weil sie möglicherweise bei kleinen Änderungen in der Hardware und der Softwareumgebung nicht mehr funktioniert. Dies wird möglicherweise nicht erkannt, außer ein leistungsfähiges Test- und Prüfungsprogramm wird kontinuierlich verwendet.

3.2.5 Bewährte Erhaltungspolicies und –abläufe einhalten

Es ist essentiell für ein OAIS, zur Erhaltung der AIPs dokumentierte Policies und Prozesse zu haben und es sollte diese auch einhalten. Insbesondere sollten AIPs niemals gelöscht werden, außer wenn es als Teil einer bewährten Policy erlaubt ist – es darf keine ad-hoc Löschungen geben. Die entsprechenden Policies und Prozesse werden zumindest abhängen von der Beschaffenheit der AIPs und von allen Absprachen zur Datensicherung, die die Archive mit anderen Archiven gemacht haben können. Zum Beispiel müssen Migrationen, die die Inhaltsinformation oder die Erhaltungsmetadaten ändern, sorgfältig überwacht und die zugehörigen Erhaltungsmetadaten komplett aktualisiert werden. Diese Detailgenauigkeit, die auch sicherstellt, dass Verarbeitungsfehler vermieden werden, erfordert, dass strenge Strategien und Prozesse vorhanden sind und dass sie auch durchgesetzt werden.

Die Produzenten und die Endnutzer-Gruppe sollten mit Übergabe- und Auslieferungs-Standards, -Policies und -Abläufen ausgestattet werden, um die Erhaltungsziele des OAIS zu unterstützen.

Die vorgesehene Zielgruppe sollte beobachtet werden, um sicherzugehen, dass die Inhaltsinformation für sie weiterhin verstehbar ist. Die vorgesehene Zielgruppe könnte das Verständnis für eine Terminologie verlieren, und die Definition der Gruppe könnte so erweitert werden, dass sie auch andere Mitglieder mit anderem Hintergrundwissen einschließt. Zum Beispiel könnte eine kontinuierliche Prüfung durch Mitglieder, die die vorgesehene Zielgruppe repräsentieren, diesen Prozess unterstützen.
Ein langfristiger Technologienutzungsplan, der aktualisiert wird, wenn sich die Technologie weiter entwickelt, ist absolut wesentlich, um zu vermeiden, dass man gezwungen ist, sehr kostspielig ein System aufrechtzuerhalten, im Notfall Systemaustausche vorzunehmen und Datenrepräsentationen kostspielig zu transformieren.

Das Archiv sollte einen formalen Nachfolgeplan, Notfall-Pläne, und/ oder Treuhandvereinbarungen haben, falls es aufgelöst wird oder der Archivträger oder der Geldgeber maßgeblich die Bedingungen ändern.

3.2.6 Die Information verfügbar machen

Per Definition macht ein OAIS die Inhaltsinformation in seinen AIPs für seine vorgesehene Zielgruppe sichtbar und verfügbar. Der Bestand kann unterschiedlich dargestellt werden, und zwar durch verschiedene Suchhilfen, die einen Überblick über die Kollektionen von AIPs geben. Einige AIPs gibt es vielleicht nur als Ergebnis von Algorithmen, die auf anderen AIPs laufen. Sie erscheinen als DIPs, die bei ihrer Auslieferung Dokumentation darüber enthalten sollten, wie sie aus anderen AIPs abgeleitet wurden. Die Erwartung der OAIS-Endnutzer in Bezug auf Zugriffsdienste kann sich von Archiv zu Archiv und auch im Lauf der Zeit, während sich die Technologie weiterentwickelt, sehr unterscheiden. Unter den gegebenen Ressourcenbeschränkungen sollte das Archiv abwägen zwischen dem Druck, einen leistungsfähigeren Zugriff zu ermöglichen und den Anforderungen der Erhaltung.

Bei einigen AIPs wird der Zugriff beschränkt sein und deshalb nur solchen Endnutzern bereitgestellt werden, die dazu berechtigt sind. Das OAIS sollte seine Zugriffspolicies und –beschränkungen veröffentlichen, damit die Rechte aller Beteiligten geschützt werden.

Grundsätzlich können Auslieferungsinformationspakete auf allen denkbaren Kommunikationswegen verbreitet werden, Netzwerke und physische Datenträger eingeschlossen.

4 Detaillierte Modelle

Der Zweck dieses Abschnitts ist es, eine detailliertere, modellhafte Sicht auf die Funktionseinheiten des OAIS und der im OAIS bearbeiteten Information zu liefern. Dies unterstützt Gestalter künftiger OAIS Systeme und stellt ein präziseres Set von Begriffen und Konzepten für die Diskussion aktueller Systeme zur Verfügung.

4.1. Funktionales Modell

Das OAIS aus Abbildung 2-1 ist in Abbildung 4-1 unterteilt in sechs Funktionseinheiten mit zugehörigen Schnittstellen. Nur die wichtigsten Informationsflüsse werden gezeigt. Die Linien, welche die Funktionseinheiten verbinden, stellen Kommunikationswege dar, über die Information in beide Richtungen fließen. Die Linien zur ‚Administration' und zur ‚Erhaltungsplanung' sind nur unterbrochen gezeichnet, um das Diagramm übersichtlicher zu machen.

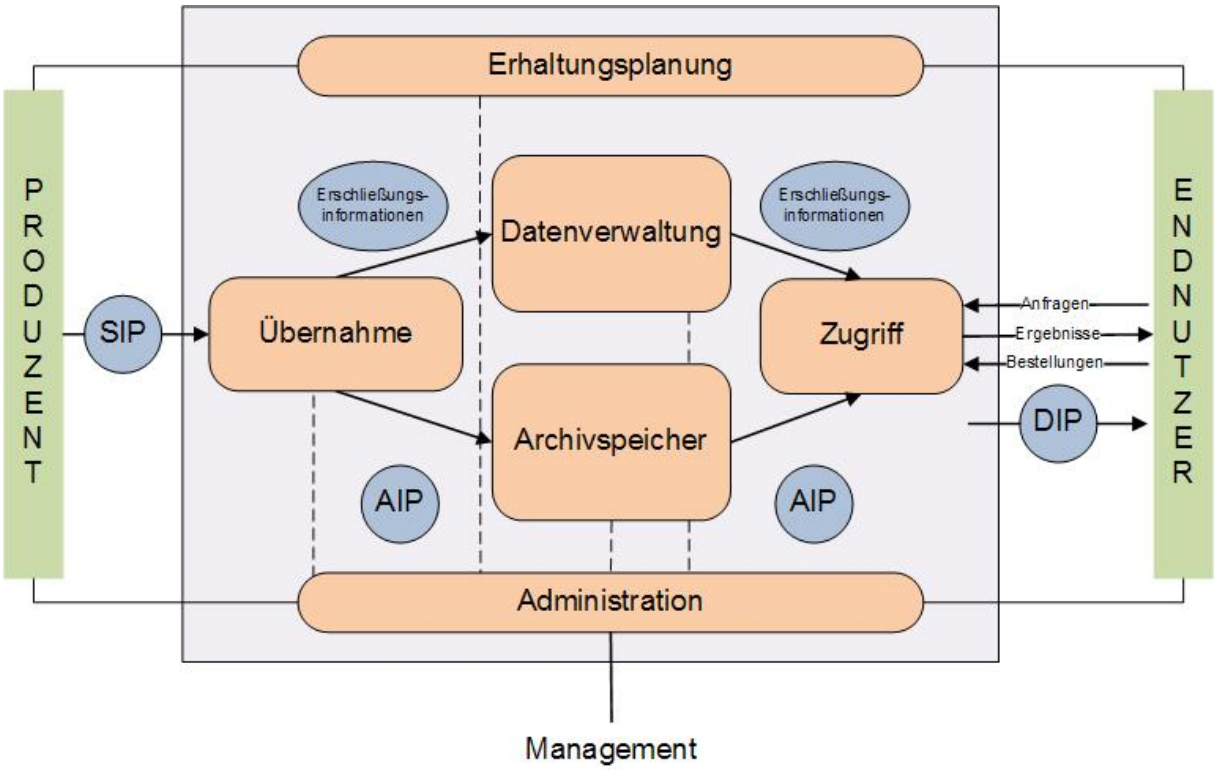

Abbildung 4-1 OAIS-Funktionseinheiten

Die Rolle, welche jede dieser Funktionseinheiten in der Abbildung 4-1 einnimmt, wird im Folgenden kurz beschrieben:

Die Funktionseinheit **‚Übernahme'** stellt Dienste und Funktionen bereit zur Annahme von Übergabeinformationspaketen (SIPs) von Produzenten (oder von internen Elementen unter Kontrolle der ‚Administration') und zur Vorbereitung für die Speicherung und Verwaltung innerhalb des Archivs. Die Übernahmefunktionen beinhalten im Einzelnen: SIPs entgegennehmen, Qualitätskontrolle an SIPs durchführen; ein Archivinformationspaket (AIP) erzeugen, das den Standards des Archivs für Datenformate und Dokumentation entspricht; Erschließungsinformationen aus den AIPs extrahieren für die Aufnahme in die Archivdatenbank und Aktualisierungen von ‚Archivspeicher' und ‚Datenverwaltung' koordinieren.

Die Funktionseinheit **‚Archivspeicher'** stellt Dienste und Funktionen bereit für die Speicherung, den Unterhalt und die Wiedergewinnung von AIPs. Die Funktionen des ‚Archivspei-

chers‘ beinhalten im Einzelnen: AIPs von der ‚Übernahme‘ entgegennehmen und in den Langzeitspeicher aufnehmen; die Speicherhierarchie verwalten; die Datenträger, auf denen Archivbestände gespeichert sind, umkopieren; routinemäßige und spezielle Fehlerkontrollen durchführen; Ressourcen für die Notfallwiederherstellung bereitstellen und AIPs an ‘Zugriff’ liefern, um Bestellungen zu erfüllen.

Die Funktionseinheit **‚Datenverwaltung‘,** stellt Dienste und Funktionen bereit für die Ergänzung, den Unterhalt und das Zugreifen sowohl auf die Erschließungsinformationen, welche die Archivbestände identifizieren und dokumentieren, als auch auf die administrativen Daten, die für die Verwaltung des Archivs benötigt werden. Die Funktionen der ‚Datenverwaltung‘ beinhalten im Einzelnen: die Funktionen des Archivdatenbanksystems (Pflege der Datenbankschemas und der Datensichten sowie der referentiellen Integrität) verwalten; Datenbankaktualisierungen durchführen (neue Erschließungsinformationen oder administrative Daten laden); Suchanfragen auf den Daten der ‚Datenverwaltung‘ ausführen, um Suchergebnisse zu erzeugen und aus diesen Suchergebnissen Berichte zu erstellen.

Die Funktionseinheit ‚**Administration‘** stellt Dienste und Funktionen bereit für den gesamten Betrieb des Archivsystems. Die Funktionen der ‚Administration‘ beinhalten im Einzelnen: Übergabevereinbarungen mit Produzenten anstoßen und aushandeln; Übergaben auf die Einhaltung der Archivstandards überprüfen und das Konfigurationsmanagement von Hard- und Software aufrechterhalten. Sie stellt außerdem Systementwicklungs-Funktionen bereit, die sowohl den Archivbetrieb überwachen und verbessern wie auch Inventare und Berichte über die Inhalte des Archivs erstellen und diese Inhalte migrieren/aktualisieren. Sie ist ebenfalls zuständig für die Festlegung und Pflege von Standards und Policies des Archivs, für die Unterstützung der Kunden und für die Aktivierung gespeicherter Anfragen.

Die Funktionseinheit **‚Erhaltungsplanung‘** stellt Dienste und Funktionen bereit zur Beobachtung des Umfeldes des OAIS und zur Erstellung von Empfehlungen und Erhaltungsplänen, mit denen sichergestellt werden soll, dass die im OAIS gespeicherte Information für die jeweilige vorgesehene Zielgruppe langfristig verfügbar und verstehbar bleiben, auch wenn die ursprüngliche Rechner-Umgebung obsolet wird. Die Funktionen der ‚Erhaltungsplanung‘ beinhalten im Einzelnen: die Inhalte des Archivs evaluieren und regelmäßig Aktualisierungen der Archivinformationen empfehlen, die Migration von Archivbeständen empfehlen, Empfehlungen für Standards und Richtlinien des Archivs entwickeln, periodische Risikoanalysen liefern und Veränderungen im technologischen Umfeld und bei den Dienstleistungsanforderungen und im Grundwissen der vorgesehenen Zielgruppe beobachten. Die ‚Erhaltungsplanung‘ gestaltet außerdem Vorlagen für Informationspakete, unterstützt und prüft die Umsetzung dieser Vorlagen in konkrete SIPs und AIPs von bestimmten Übergaben. Die ‚Erhaltungsplanung‘ entwickelt ebenfalls detaillierte Migrationspläne, Software-Prototypen und Testpläne, um die Umsetzung der Migrationsziele der ‚Administration‘ zu ermöglichen.

Die Funktionseinheit **‚Zugriff‘** stellt Dienste und Funktionen bereit, welche die Endnutzer dabei unterstützen, das Vorhandensein, die Beschreibung, den Standort und die Verfügbarkeit von Information, die im OAIS gespeichert sind, zu ermitteln, und die den Endnutzern ermöglichen, Informationsprodukte anzufragen und entgegenzunehmen. Die Funktionen des ‚Zugriffs‘ beinhalten: mit Endnutzern kommunizieren, um Anfragen entgegenzunehmen; Zugriffskontrollen anwenden, um den Zugang zu speziell geschützten Information zu beschränken; die Bearbeitung von Anfragen bis zum erfolgreichen Abschluss koordinieren; Antworten erzeugen (Auslieferungsinformationspakete, Suchergebnisse, Berichte) und Antworten an Endnutzer ausliefern.

Zusätzlich zu den oben beschriebenen Funktionseinheiten wird eine Reihe von **Allgemeinen Diensten** vorausgesetzt. Diese Dienste bilden eine weitere Funktionseinheit in diesem Modell. Diese Funktionseinheit ist jedoch so grundlegend und allgegenwärtig, dass sie um der Klarheit willen in Abbildung 4-1 nicht dargestellt ist.

4.1.1 Detaillierte Beschreibung der Funktionseinheiten

In den folgenden Unterabschnitten werden spezifische Informationsströme zwischen den Funktionseinheiten *kursiv* dargestellt, wenn sie zum ersten Mal im Text erscheinen. Die detaillierten Beschreibungen in den Unterabschnitten werden durch Diagramme ergänzt (Abbildungen 4-2 bis 4-7), die lediglich die Hauptdatenflüsse innerhalb und zwischen den Funktionseinheiten zeigen. Aus Gründen der Übersichtlichkeit werden unwichtigere Datenflüsse wie bspw. Bestätigungsmeldungen weggelassen. Anhang A[14] enthält eine Abbildung, in welcher die Abbildungen 4-2 bis 4-7 kombiniert sind, um den konsistenten Gesamtzusammenhang zu zeigen. Allerdings soll dies nicht als Empfehlung für eine Gestaltung bzw. eine Umsetzung verstanden werden, und von konkreten Umsetzungen soll nicht erwartet werden, dass sie die dargestellten Funktionen eins zu eins abbilden; sie können beispielsweise einzelne Funktionen kombinieren oder die Funktionalität auf eine andere Weise aufteilen.

4.1.1.1 Allgemeine Dienste

Moderne, verteilte Rechneranwendungen setzen eine Vielzahl von unterstützenden Diensten wie bspw. Kommunikation zwischen Prozessen, Namensdienste, temporäre Speicherzuweisung, Ausnahmenbehandlung, Sicherheit, Datensicherung und Verzeichnisdienste voraus. Viel hervorragende Arbeit wurde bereits auf dem Gebiet der Referenzmodelle für offene Systemumgebungen geleistet. Beispiele solcher Dienste sind:

Betriebssystem-Dienste stellen die Grundleistungen bereit, die für die Bedienung und Verwaltung einer Rechnerplattform benötigt werden, und bieten eine Schnittstelle zwischen der Anwendungssoftware und der Plattform. Diese Grundleistungen beinhalten:

- Kernoperationen bieten Basis-Dienste, die nötig sind, um Prozesse zu erzeugen und zu verwalten, Programme auszuführen, Signale zu definieren und zu kommunizieren, Systemuhr-Operationen zu definieren und auszuführen, Dateien und Datenverzeichnisse zu verwalten sowie Ein- und Ausgabeprozesse von und zur externen Umgebung zu kontrollieren.

- Befehle und Dienstprogramme beinhalten Mechanismen für Operationen auf Operatorenebene, wie das Vergleichen, Drucken und Darstellen von Dateiinhalten; das Editieren von Dateien; die musterbasierte Suche; das Ausdrücke auswerten; das Protokollieren von Nachrichten; die Verschiebung von Dateien zwischen Datenverzeichnissen; das Sortieren von Daten; das Abarbeiten von Befehlssequenzen; und das Zugreifen auf Umgebungsinformationen.

- Die Echtzeit-Erweiterung beinhaltet die Anwendungs- und Betriebssystem-Schnittstellen, die benötigt werden, um jene Anwendungsbereiche zu unterstützen, welche eine deterministische Ausführung, Verarbeitung und Rückmeldung erfordern. Die Erweiterung bestimmt die Anwendungs-Schnittstelle zu grundlegenden Systemdiensten für Eingabe/Ausgabe, den Zugriff auf das Dateisystem sowie die Prozessverwaltung.

- Die Systemverwaltung beinhaltet Möglichkeiten, Zugriff und Zuteilung von Ressourcen an Benutzer zu definieren und zu verwalten (d.h., welche Ressourcen verwaltet werden und welche Zugriffskategorien es gibt) sowie die Konfigurations- und Leistungsverwaltung von Geräten, Dateisystemen, Verwaltungsprozessen (Auftragsabrechnung), Warteschlangen, Maschinen-/Plattformprofilen, Autorisierung von Ressourcennutzung und Systemsicherung zu definieren und zu verwalten.

[14] Die Anhänge sind in der deutschen Übersetzung nicht enthalten.

- Die Sicherheitsdienste des Betriebssystems spezifizieren, wie der Zugriff von Benutzern und Benutzungsprozessen auf Systemdaten, Funktionen, Hardware und Softwareressourcen kontrolliert wird.

Netzwerkdienste stellen die Möglichkeiten und Mechanismen zur Verfügung, um verteilte Anwendungen, die Zugriff auf Daten benötigen, und die Interoperabilität von Anwendungen in heterogenen, vernetzten Umgebungen zu unterstützen. Diese Dienste beinhalten im Einzelnen:

- Die Datenkommunikation beinhaltet Programmierschnittstellen (API) und Protokollspezifikationen für eine zuverlässige, transparente, und durchgehende Datenübertragung über Kommunikationsnetzwerke.

- Der transparente Datei-Zugriff bietet Zugriff auf verfügbare Dateien, die sich an irgendeiner Stelle in einem heterogenen Netzwerk befinden.

- Der Personal/Mikrocomputer-Support bietet Unterstützung für die Interoperabilität mit Systemen, die auf anderen Betriebssystemen basieren, insbesondere auf Betriebssystemen für Mikrocomputer, die möglicherweise nicht in einem nationalen oder internationalen Standard formal spezifiziert sind.

- Die Remote Procedure Call-Dienste enthalten Spezifikationen, die erlauben, den lokalen Aufruf von Prozessen auf eine verteilte Umgebung auszuweiten.

- Die Dienste für die Netzwerksicherheit beinhalten Kontrollen über den Zugriff, die Authentifizierung, die Vertraulichkeit, die Integrität, die Nichtabstreitbarkeit und die Verwaltung der Kommunikation zwischen Sendern und Empfängern von Information in einem Netzwerk.

Sicherheitsdienste stellen Möglichkeiten und Mechanismen bereit, um sensible Information und Bearbeitungen im Informationssystem zu schützen. Die angemessene Schutzstufe wird auf der Grundlage des Werts der Information für die End-Benutzer der Anwendung und der wahrgenommenen Bedrohungen festgelegt. Diese Dienste beinhalten im Einzelnen:

- Der Identifizierungs-/Authentifizierungsdienst bestätigt die Identitäten von Personen, welche die Ressourcen des Informationssystems nutzen wollen. Zusätzlich kann die Authentifizierung auch auf die Lieferanten von Daten angewendet werden. Die Authentifizierung kann am Anfang oder während einer Sitzung erfolgen.

- Die Zugriffskontroll-Dienste verhindern die unerlaubte Nutzung von Ressourcen des Informationssystems. Außerdem verhindern sie nicht erlaubte Arten der Nutzung einer Ressource. Dieser Dienst kann auf verschiedene Aspekte des Zugriffs auf eine Ressource angewendet werden (z.B. Zugriff auf Verbindungen zu einer Ressource, Lesen, Schreiben oder Löschen einer Informations-/Datenressource oder Ausführung einer Verarbeitungsressource) oder auf jeglichen Zugriff auf eine Ressource.

- Die Dienste zur Datenintegrität stellen sicher, dass Daten nicht unautorisiert verändert oder zerstört werden. Diese Dienste werden angewendet auf Daten in dauerhaften Datenspeichern und beim Austausch von Nachrichten.

- Die Dienste zum Schutz der Vertraulichkeit stellen sicher, dass Daten nur für berechtigte Personen oder Computerprozesse zugänglich oder bekannt gemacht werden. Diese Dienste werden auf Geräte angewendet, welche die Interaktion zwischen

Menschen und dem Informationssystem erlauben. Zudem werden diese Dienste dafür sorgen, dass das Beobachten von Nutzungsmustern der Kommunikationsressourcen nicht möglich ist.

- Die Dienste zur Nichtabstreitbarkeit stellen sicher, dass niemand, der an einem Informationsaustausch beteiligt ist, seine Beteiligung leugnen kann. Dieser Dienst kann eine oder beide von zwei möglichen Formen annehmen. Erstens erhält der Empfänger von Daten einen Herkunftsnachweis zu diesen Daten. Dies schützt gegen jeden Versuch des Absenders, die Versendung von Daten oder ihrer Inhalte fälschlicherweise abzustreiten. Zweitens erhält der Sender der Daten einen Nachweis für die Lieferung der Daten. Dies schützt vor jedem späteren Versuch des Empfängers, den Empfang von Daten oder ihrer Inhalte fälschlicherweise abzustreiten.

4.1.1.2 Übernahme

Die Funktionen der Funktionseinheit ‚Übernahme' sind in Abbildung 4-2 illustriert.

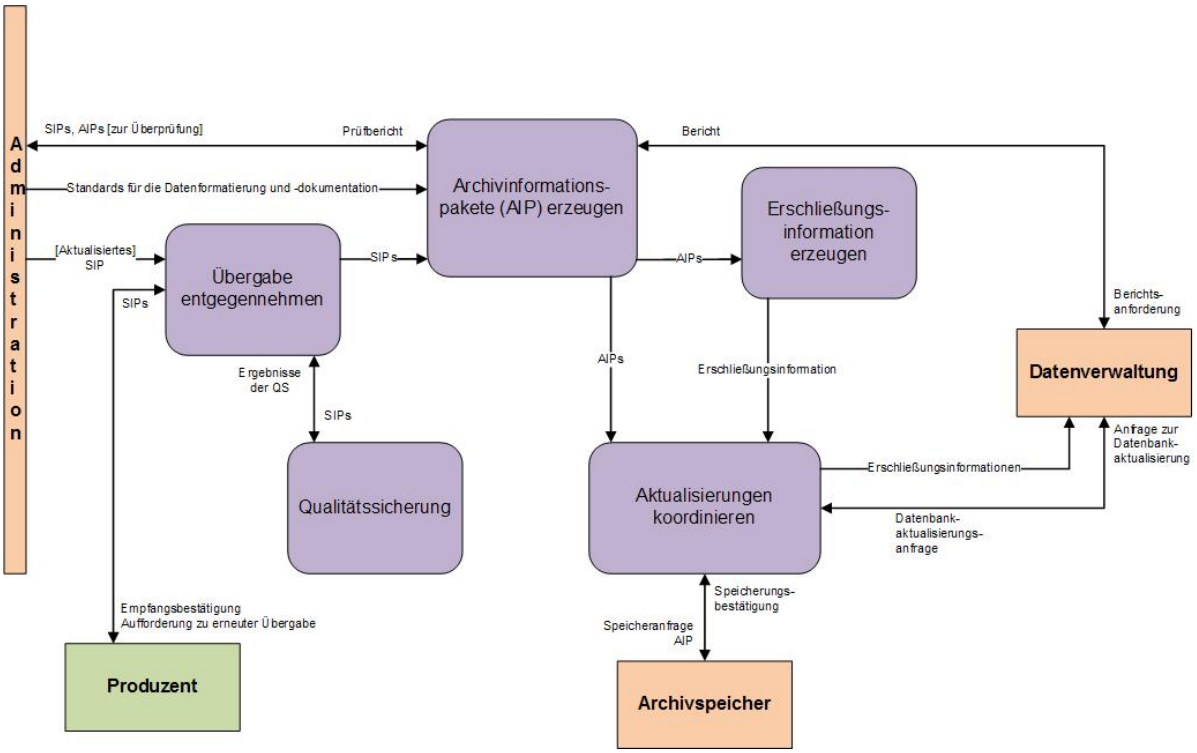

Abbildung 4-2: Funktionen der Funktionseinheit ‚Übernahme'

Die Funktion **‚Übergabe entgegennehmen**' stellt angemessene Speichermöglichkeiten oder –geräte zur Verfügung, um ein Übergabeinformationspaket (SIP) vom Produzenten (oder von der ‚Administration') entgegenzunehmen. Digitale SIPs können entweder über einen elektronischen Transfer abgeliefert werden (z.B. FTP), von Datenträgern kopiert werden, die dem Archiv übergeben wurden, oder einfach im Dateiverwaltungssystem des Archivs für den Zugriff geladen werden (z.B. CD-ROM). Nicht-digitale SIPs werden wahrscheinlich auf konventionellen Wegen abgeliefert. Die Funktion ‚Übergabe entgegennehmen' kann einen rechtserheblichen Transfer der Obhut über die Inhaltsinformationen im SIP darstellen und spezielle Zugriffskontrollen auf diese Inhalte nötig machen. Diese Funktion liefert eine *Empfangsbestätigung* für das SIP an den Produzenten, die im Fall eines Fehlers in der SIP-Übergabe eine *Aufforderung zu erneuter Übergabe* beinhalten kann.

Evidenz für die Authentizität wird vom Produzenten als Teil der Erhaltungsmetadaten in der Übergabe geliefert, und diese Evidenz wird über die Zeit hinweg durch das Archiv erhalten, aktualisiert und/oder angereichert. Der Produzent kann als Bestandteil der Provenienzinformation **Informationseigenschaftsbeschreibungen** von **Informationseigenschaften** liefern, die über die Zeit hinweg erhalten werden sollen, und freilich auch von solchen Informationseigenschaften, deren Erhaltung nicht nötig ist. Das Archiv kann diese auch selber definieren. Eine Informationseigenschaft ist derjenige Teil der Inhaltsinformation, der durch die Informationseigenschaftsbeschreibung beschrieben wird. Eine Informationseigenschaftsbeschreibung beschreibt einen Teil des Informationsinhaltes eines Inhaltsinformationsobjektes, der für einen bestimmten Zweck hervorgehoben wird. Der genaue Ausdruck oder Wert dieses Teils des Informationsinhalts wird über die entsprechenden Teile des Inhaltsdatenobjekts und dessen Repräsentationsinformation vermittelt. Das Archiv kann eine Beurteilung der Authentizität seiner Bestände abgeben, basierend auf Praxis und Empfehlungen der Fachgemeinschaft (einschließlich von bester Praxis, Richtlinien, Standards und rechtlichen Anforderungen). Wissenschaftliche Archive können zum Beispiel weniger strenge Beurteilungskriterien haben als staatliche Archive: indes, der Endnutzer kann sich sein eigenes Urteil über die Authentizität bilden, ausgehend von der Evidenz, welche ihm die Erhaltungsmetadaten liefern.

Die Funktion der **‚Qualitätssicherung'** bestätigt *(Ergebnisse der QS)* den erfolgreichen Transfer des *SIP*s in den Zwischenspeicher. Für digitale Übergaben können zu diesen Mechanismen zudem zyklische Redundanzprüfungen (CRCs) oder Prüfsummen kommen, die mit jeder einzelnen Datei verknüpft sind, oder es können Systemprotokoll-Dateien verwendet werden, um jegliche Fehler beim Datentransfer oder beim Lesen und Schreiben von Datenträgern aufzuzeichnen und zu identifizieren.

Die Funktion **‚Archivinformationspaket (AIP) erzeugen'** transformiert ein oder mehrere *SIPs* in ein oder mehrere *AIPs,* welche konform zu den *Standards für Datenformatierung und Dokumentation* des Archivs sind. Dies kann eine Konvertierung der Dateiformate und der Repräsentationsform der Daten oder die Reorganisation der Inhaltsinformation in den SIPs umfassen. Die Funktion ‚AIP erzeugen' kann *Berichtsanforderungen* an die ‚Datenverwaltung' senden, um *Berichte* mit Information zu erhalten, die von ihr dazu gebraucht werden, die Erschließungsinformationen zu produzieren, die das AIP vervollständigen. Die Funktion sendet außerdem *SIPs* oder *AIPs* zur *Überprüfung* an die Funktion ‚Übergabe prüfen' innerhalb der ‚Administration' und erhält einen *Prüfbericht* zurück.

Die Funktion **‚Erschließungsinformationen erzeugen'** extrahiert Erschließungsinformationen aus den *AIPs* und sammelt *Erschließungsinformationen* aus anderen Quellen, um sie an ‚Aktualisierungen koordinieren' und schließlich an die ‚Datenverwaltung' zu liefern. Dies betrifft Metadaten, welche die Suche und das Wiederauffinden von AIPs unterstützen (z.B. wer, was, wann, wo, warum), kann aber auch spezielle Browserprodukte für die Verwendung in Findmitteln einschliessen (Thumbnails, Bilder).

Die Funktion **‚Aktualisierungen koordinieren'** ist für den Transfer der *AIPs* in den ‚Archivspeicher' und der Erschließungsinformationen in die ‚Datenverwaltung' verantwortlich. Der Transfer der *AIPs* schließt eine *Speicherungsanfrage* mit ein und kann in Form eines elektronischen, physischen oder virtuellen (d.h. die Daten bleiben an ihrem Platz) Transfers erfolgen. Nachdem der Transfer abgeschlossen und überprüft ist, sendet der ‚Archivspeicher' eine *Speicherungsbestätigung* zurück, welche die Speicheridentifikation des AIPs anzeigt (oder verifiziert). Die Funktion ‚Aktualisierungen koordinieren' integriert zudem die Speicheridentifikation in die Erschließungsinformationen zum AIP und übermittelt diese im Rahmen einer *Datenbankaktualisierungsanfrage* an die ‚Datenverwaltung'. Im Gegenzug liefert die

‚Datenverwaltung' eine *Rückmeldung zur Datenbankaktualisierung,* die den Status der Aktualisierung angibt. Aktualisierungen der ‚Datenverwaltung' können ohne einen entsprechenden Transfer in den ‚Archivspeicher' stattfinden, wenn das SIP Erschließungsinformationen für ein AIP enthält, das sich bereits im ‚Archivspeicher' befindet.

4.1.1.3. Archivspeicher

Die Funktionen der Funktionseinheit ‚Archivspeicher' sind in Abbildung 4-3 dargestellt. Der Begriff ‚Speichermedium' wird benutzt, um einen oder mehrere Mechanismen für die lokale oder entfernte Speicherung digital kodierter Information zu bezeichnen.

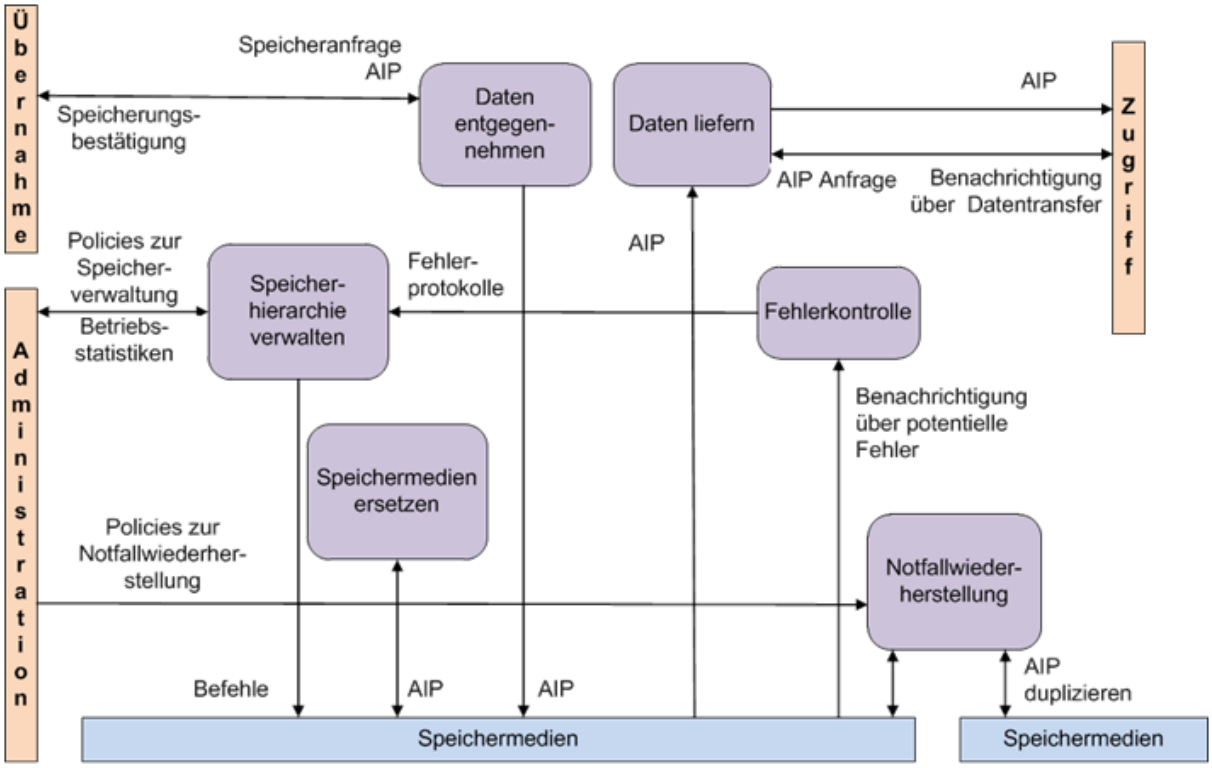

Abbildung 4-3: Funktionen der Funktionseinheit ‚Archivspeicher'

Die Funktion ‚**Daten entgegennehmen'** empfängt eine *Speicheranfrage* und ein *AIP* aus der ‚Übernahme' und verschiebt das *AIP* in den permanenten Speicher innerhalb des Archivs. Die Transferanfrage muss unter Umständen die erwartete Nutzungsfrequenz der Datenobjekte, welche das AIP bilden, angeben, damit die geeignetsten Speichergeräte oder Speichermedien für die Speicherung des AIP gewählt werden können. Diese Funktion wählt den Typ des Speichermediums, bereitet die Geräte oder Datenträger vor und führt den physischen Transfer auf die Archivspeichermedien durch. Nach Abschluss des Transfers sendet diese Funktion eine *Speicherungsbestätigung*, welche die Speicheridentifikation der AIPs enthält, an die ‚Übernahme'.

Die Funktion **‚Speicherhierarchie verwalten'** legt über *Befehle* die Inhalte der AIPs auf geeigneten Speichermedien ab und stützt sich dabei auf die *Policies zur Speicherverwaltung,* die Betriebsstatistiken oder auf Anweisungen in der Speicheranfrage aus der ‚Übernahme'. Sie erfüllt spezielle Leistungen, die für das AIP verlangt werden, und notwendige spezielle Sicherheitsanforderungen. Sie gewährleistet die angemessene Schutzstufe für das AIP. Diese Leistungen umfassen Online-, Offline- oder Nearline-Speicherung, die benötigte

Datendurchsatzrate, die maximal erlaubte Bit-Fehler-Rate oder spezielle Behandlungs- oder Datensicherungsverfahren. Die Funktion überwacht *Fehlerprotokolle*, um sicherzustellen, dass die AIPs während der Transfers nicht beschädigt werden. Außerdem liefert diese Funktion der ‚Administration' *Betriebsstatistiken*, welche eine Übersicht bieten über die vorhandenen Speichermedien, die verfügbaren Speicherkapazitäten auf den verschiedenen Stufen der Speicherhierarchie und die Nutzungsziffern.

Die Funktion **‚Speichermedien ersetzen'** ermöglicht die Reproduktion der *AIPs* über die Zeit hinweg. Im Rahmen der Funktion ‚Speichermedien ersetzen' dürfen Inhaltsinformationen und Erhaltungsmetadaten nicht verändert werden. Jedoch können die Daten, welche die Verpackungsinformation bilden, verändert werden, solange sie dieselben Funktionen leisten und direkt angewendet werden können ohne einen Informationsverlust zu verursachen. Die Migrationsstrategie wählt ein Speichermedium unter Berücksichtigung der zu erwartenden und der aktuellen Fehlerrate, der Leistungsfähigkeit und der anfallenden Anschaffungs- und Betriebskosten der verschiedenen Typen von Speichermedien. Wenn vom Speichermedium abhängige Eigenschaften (z.B. Bandblockgrößen, CD-ROM Volume Informationen) in die Inhaltsinformationen aufgenommen worden sind, so muss ein Weg gefunden werden, diese Information zu erhalten, wenn auf Speichermedien mit einer höheren Kapazität und einer anderen Speicherarchitektur migriert wird. Unter Vorausnahme der Terminologie von 5.1.3 kann diese Funktion eins zu eins ‚auffrischen', ‚replizieren' und ‚umverpacken'. Ein Beispiel für ‚umverpacken' ist die Migration auf neue Speichermedien unter einem neuen Betriebs- und Dateisystem, wobei die Inhaltsinformationen und die Erhaltungsmetadaten unabhängig von den Dateisystemen sind. Allerdings müssen komplexe ‚Umverpackungen' und alle ‚Transformationen' unter der Kontrolle der ‚Administration' über die Funktion ‚Archivinformationen aktualisieren' ausgeführt werden, damit die Erhaltung der Information zu gewährleistet werden kann. (Vgl. 5.1.3 für eine detaillierte Beschreibung der Migration.)

Die Funktion **‚Fehlerkontrolle'** gibt eine statistisch akzeptable Gewissheit, dass keine Komponenten des AIPs während eines internen Datentransfers in den ‚Archivspeicher' beschädigt werden. Diese Funktion setzt voraus, dass die gesamte Hard- und Software innerhalb des Archivs *Benachrichtigungen über potenzielle Fehler* liefert und dass diese in standardisierten *Fehlerprotokollen* zusammengeführt und vom Personal des ‚Archivspeichers' überprüft werden. Die Persistenzinformation der Erhaltungsmetadaten liefert einige Gewissheit, dass die Inhaltsinformationen nicht verändert worden sind, wenn das AIP verschoben oder benutzt wird. Entsprechende Information wird benötigt, um die Erhaltungsmetadaten selber zu schützen. Außerdem kann ein Standardmechanismus für die Nachverfolgung und Überprüfung der Gültigkeit aller Datenobjekte im Archiv eingesetzt werden. Zum Beispiel können zyklische Redundanzprüfungen (CRCs) für jede einzelne Datei regelmäßig durchgeführt werden. Es kann auch eine höhere Stufe erreicht werden, wie zum Beispiel durch die Reed-Solomon-Kodierung, um Fehlersuche und Fehlerkorrektur zu kombinieren. Die Verfahren der Speicher-Einrichtung sollten eine stichprobenartige Überprüfung der Integrität der Datenobjekte mit Hilfe von zyklischen Redundanzkontrollen (CRC) oder anderen Mechanismen bereitstellen.

Die Funktion **‚Notfallwiederherstellung'** stellt einen Mechanismus zur Verfügung, der die digitalen Inhalte der Archivkollektion dupliziert und diese Duplikate bspw. in einer physisch getrennten Anlage aufbewahrt. Diese Funktion wird normalerweise ausgeführt, indem die Inhalte des Archivs auf entfernbare Speichermedien kopiert werden (z.B. DLT-Magnetbänder, CD-ROM), kann aber auch durch einen Transport von Hardware oder den Datentransfer über ein Netzwerk stattfinden. Die Details werden durch die ‚Administration' in den *Policies zur Notfallwiederherstellung* spezifiziert.

Die Funktion **‚Daten liefern'** stellt dem ‚Zugriff' Kopien von gespeicherten *AIPs* zur Verfügung. Diese Funktion erhält eine *AIP Anfrage*, welche das/die verlangte/n AIP/s identifiziert, und stellt diese auf dem verlangten Datenträgertyp bereit oder verschiebt sie in einen Zwischenspeicher. Diese Funktion sendet außerdem eine *Benachrichtigung über den Datentransfer* an den ‚Zugriff', sobald eine Anfrage abgeschlossen worden ist.

4.1.1.4 Datenverwaltung

Die Funktionen der Funktionseinheit ‚Datenverwaltung' sind in Abbildung 4-4 dargestellt.

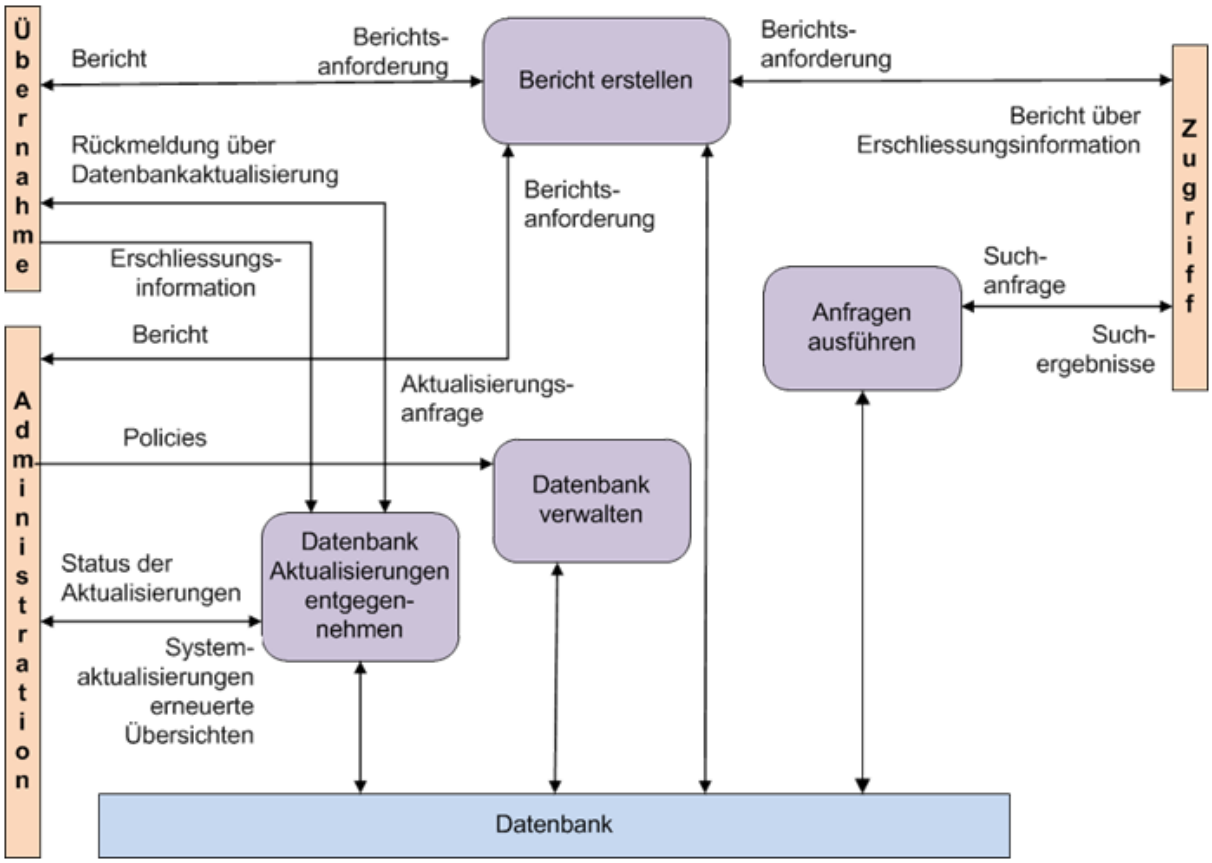

Abbildung 4-4: Funktionen der Funktionseinheit ‚Datenverwaltung'

Die Funktion **‚Datenbank verwalten'** ist für die Integrität der Datenbank der ‚Datenverwaltung' verantwortlich, welche einen Speichermechanismus für Erschließungs- und Systeminformationen bereitstellt, der auf eine bestimmte Weise abgefragt werden kann. Erschließungsinformationen identifizieren und beschreiben die Archivbestände und die Systeminformationen dienen dazu, den Betrieb des Archivs zu unterstützen. Die Funktion ‚Datenbank verwalten' ist verantwortlich für die Erstellung aller Schemata oder Tabellen, die für die Unterstützung der Funktionen der ‚Datenverwaltung' benötigt werden. Sie ist ebenso verantwortlich für die Möglichkeit, benutzerspezifische Sichten auf die Inhalte dieses Speichers zu erstellen, zu pflegen und zugänglich zu machen sowie für die interne Validierung (z.B. referentielle Integrität) der Inhalte der Datenbank. Die Funktion ‚Datenbank verwalten' wird gemäss den Policies der ‚Administration' ausgeführt.

Die Funktion **‚Anfragen ausführen'** bekommt eine *Suchanfrage* vom ‚Zugriff' und bearbeitet die Anfrage, um ein *Suchergebnis* zu erzeugen, welches an den Anfragenden weitergeleitet wird.

Die Funktion **‚Bericht erstellen'** empfängt eine *Berichtsanforderung* von der ‚Übernahme', vom ‚Zugriff' oder von der ‚Administration' und führt alle Abfragen oder Prozesse durch, die nötig sind, um den *Bericht* zu generieren, der an den Anfragenden geliefert wird. Typische Berichte können Übersichten der Archivbestände nach Kategorien beinhalten oder Benutzungsstatistiken über die Zugriffe auf Archivbestände. Die Funktion kann auch eine Berichtsanforderung vom ‚Zugriff' erhalten und *Erschließungsinformationen* für ein bestimmtes AIP bereitstellen.

Die Funktion **‚Datenbankaktualisierungen entgegennehmen'** ergänzt, modifiziert oder löscht Information im persistenten Speicher der ‚Datenverwaltung'. Die Hauptquelle der Aktualisierungen ist die ‚Übernahme', welche *Erschließungsinformationen* für die neuen AIPs liefert, sowie die ‚Administration', die *Systemaktualisierungen* und *erneuerte Übersichten*[15] liefert. Transaktionen von der ‚Übernahme' bestehen aus Erschließungsinformationen, die neu im Archiv gespeicherte AIPs identifizieren. Systemaktualisierungen enthalten alle auf das System bezogenen Informationen (Betriebsstatistiken, Informationen über Endnutzer und Anfragestatus). Erneuerte Übersichten werden durch periodische Überprüfungen und Aktualisierungen von Informationswerten (z.B. Kontaktnamen, Adressen, Zugangskontrolle und Berechtigungspolicies) erzeugt. Die Funktion ‚Datenbankaktualisierungen entgegennehmen' liefert der ‚Administration' regelmäßig zusammenfassende Berichte über den *Status von Aktualisierungen* in der Datenbank und sendet ebenfalls eine *Rückmeldung über Datenbankaktualisierungen* an die ‚Übernahme'.

[15] Im Original: "Review Updates"

4.1.1.5 Administration

Die Funktionen der Funktionseinheit ‚Administration' sind in Abbildung 4-5 dargestellt. Wie weiter oben festgehalten (siehe Seite 4-3) werden einzelne Aktivitäten, die in diesem Abschnitt im Detail dargestellt sind, nicht auf alle konkreten Implementationen anwendbar sein; bei den Kunden-Diensten bspw. ist die Aktivität Abrechnung nur dort vorhanden, wo sie wirklich benötigt wird.

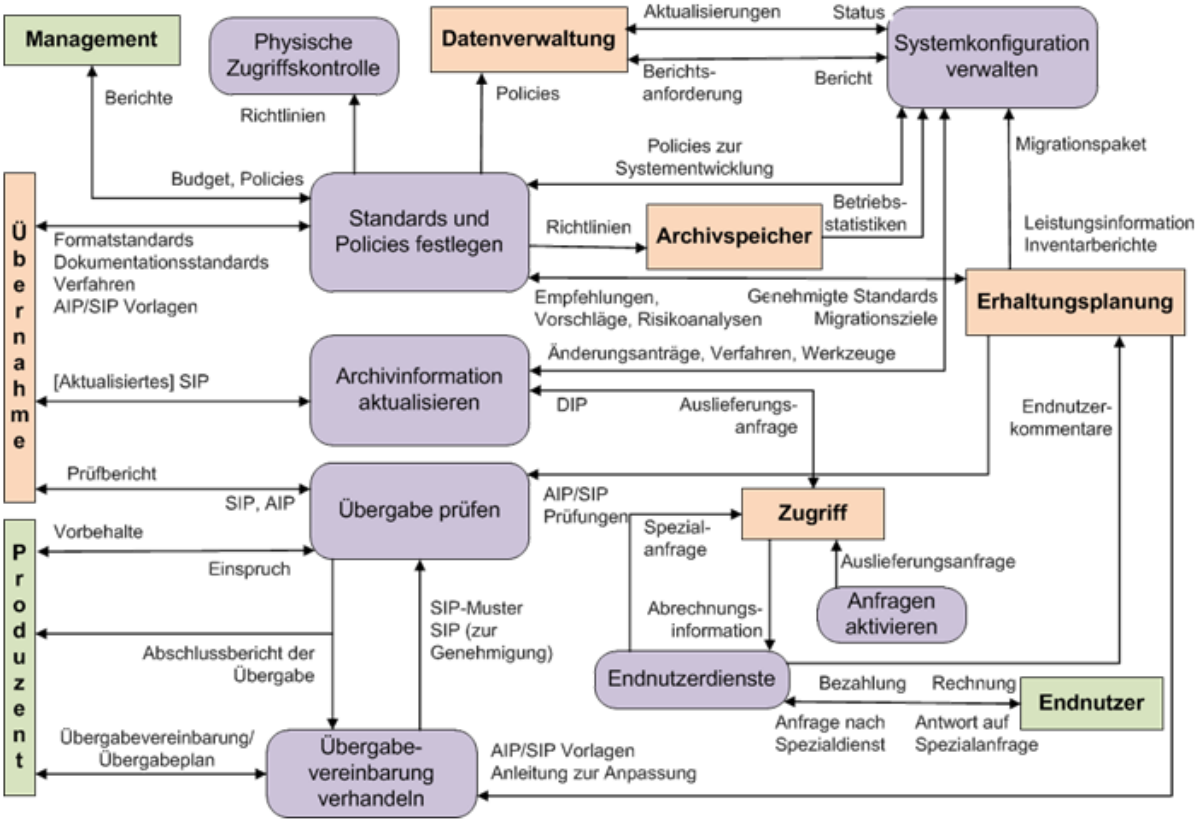

Abbildung 4-5: Funktionen der Funktionseinheit ‚Administration'

Die Funktion ‚**Übergabevereinbarung verhandeln'** bemüht sich um erstrebenswerte Archivinformationen für das OAIS und handelt mit den Produzenten *Übergabe-vereinbarungen* aus. Diese Funktion vereinbart ebenfalls einen *Datenübergabeplan* mit dem Produzenten. Sie unterhält einen Zeitplan mit den erwarteten Datenübergabesitzungen, die notwendig sind, um ein oder mehrere SIPs in das OAIS zu transferieren, und weist die für ihre Übernahme benötigten Ressourcen nach. Diese Funktion bekommt von der ‚Erhaltungsplanung' *AIP/SIP-Vorlagen* sowie *Anleitungen zur Anpassung* und sendet *SIP-Muster* und *SIPs* zur Funktion ‚Übergabe prüfen' als Bestandteil des Genehmigungsprozesses der Übergabe. Die Datenübergabeformate und -verfahren müssen in den Datenübergaberichtlinien des Archivs klar dokumentiert sein und die zu liefernden Elemente müssen durch den Produzenten in der Übergabevereinbarung genau bestimmt werden. Der Producer-Archive Interface Methodology Abstract Standard (ISO 20652:2006) ist hier anwendbar, ebenso Teile von ISO 15489-1:2001 und ISO/TR 15489-2:2001 und andere ISO Standards.

Die Funktion **‚Systemkonfiguration verwalten'** stellt Systemanalysen für das Archivsystem bereit, um kontinuierlich die Funktionalität des gesamten Archivsystems zu überwachen und systematisch Veränderungen in der Konfiguration zu kontrollieren. Diese Funktion bewahrt Integrität und Lenkbarkeit der Konfiguration während sämtlicher Phasen im Lebenszyklus eines Systems. Sie überprüft außerdem den Systembetrieb sowie die Leistung und die Nutzung des Systems. Sie sendet *Berichtsanforderungen* über Systeminformationen an die

‚Datenverwaltung' und erhält *Berichte* zurück. Sie bekommt *Betriebsstatistiken* vom ‚Archivspeicher'. Sie fasst diese Berichte zusammen und liefert der ‚Erhaltungsplanung' regelmäßig OAIS *Leistungsinformationen* und *Inventarlisten* der Archivbestände. Sie erhält *Migrationspakete* von der ‚Erhaltungsplanung'. Sie erhält *Policies zur Systementwicklung* aus der Funktion ‚Standards und Richtlinien festlegen'. Auf Basis dieser Eingänge entwickelt und implementiert sie Pläne zur Weiterentwicklung des Systems. Sie sendet *Änderungsanträge, Verfahren* und *Werkzeuge* an ‚Archivinformation aktualisieren'.

Die Funktion **‚Archivinformation aktualisieren'** stellt einen Mechanismus zur Aktualisierung der Inhalte des Archivs bereit. Sie bekommt *Änderungsanträge, Verfahren* und *Werkzeuge* aus ‚Systemkonfiguration verwalten'. Sie stellt Aktualisierungen zur Verfügung, indem sie eine *Auslieferungsanfrage* an den ‚Zugriff' sendet, die Inhalte der darauf erhaltenen *DIPs* aktualisiert und sie als *SIPs* neu an die ‚Übernahme' schickt.

Die Funktion **‚Physische Zugriffskontrolle'** stellt Mechanismen zur Verfügung, um den physischen Zugang (Türen, Schlösser, Wächter) zu Bestandteilen des Archivs gemäß den Richtlinien des Archivs zu beschränken oder zu erlauben.

Die Funktion **‚Standards und Policies festlegen'** ist verantwortlich für das Erlassen und die Pflege der Standards und Policies des Archivsystems. Sie erhält vom Management Informationen zum *Budget* und *Policies,* wie z.B. die Satzung des OAIS, den Zuständigkeitsbereich, Leitlinien zur Verwendung der Mittel und Policies zur Preisbildung. Sie liefert dem Management regelmäßig *Berichte* ab. Sie erhält von der ‚Erhaltungsplanung' *Empfehlungen* zur Verbesserung des Archivsystems, *Vorschläge* für neue Archivdatenstandards und periodische *Risikoanalysen*. Sie muss Risiken durch unvorhergesehene Ereignisse (ungeplante Betriebsunterbrüche aufgrund von Netzausfällen, Software-Fehler, Hardwaredefekte, menschliche Fehlleistungen, Festplatten-Absturz etc.) in Betracht ziehen und angemessene Maßnahmen ergreifen um das Risiko, dass das Archiv seinen Verpflichtungen nicht nachkommen kann, zu minimieren. Sie bekommt außerdem Leistungsinformationen und Inventare der Archivbestände von ‚Systemkonfiguration verwalten'. Auf Basis dieser Eingänge und Analysen werden Archivstandards und *Policies* festgelegt und zur Umsetzung an andere Funktionen der ‚Administration' und an andere Funktionseinheiten geschickt. Die Standards umfassen *Formatstandards, Dokumentationsstandards* und *Verfahren,* die während dem Übernahmeprozess eingehalten werden sollen. Die Funktion liefert der ‚Erhaltungsplanung' *genehmigte Standards* und *Migrationsziele*. Sie entwickelt außerdem sowohl *Policies zur Speicherverwaltung* (für die Archivspeicherhierarchie), einschließlich von Migrationsrichtlinien, um sicher zu stellen, dass Archivspeicherformate nicht obsolet werden, als auch Policies zur Datenbankverwaltung. Sie entwickelt *Policies zur Notfallwiederherstellung*. Sie legt außerdem *Policies* zur Sicherheit der Archivbestände fest, welche auch die physische Zugriffskontrolle, wie bspw. DRM, und die Anwendung von Techniken der Fehlerkontrolle im gesamten Archiv einschließen.

Die Funktion **‚Übergabe prüfen'** kontrolliert, ob Übergaben *(SIP* oder *AIP)* den Spezifikationen der Übergabevereinbarung entsprechen. Diese Funktion bekommt von der ‚Erhaltungsplanung' *AIP/SIP-Prüfungen;* sie kann auch einen externen Ausschuss dazu einbeziehen (z.B. wissenschaftliche und technische Prüfung). Diese Überprüfungen stellen fest, ob die AIP/SIP-Vorlagen in der ‚Übernahme' korrekt angewendet worden sind. Die Funktion muss verifizieren, ob die Qualität der Daten den Anforderungen des Archivs und des Prüfausschusses genügt. Sie muss verifizieren, dass adäquate Repräsentationsinformationen und Erhaltungsmetadaten vorhanden sind, um sicherzugehen, dass die Inhaltsinformation für die vorgesehene Zielgruppe verstehbar und unabhängig nutzbar ist. Die Form der Überprüfung kann in Abhängigkeit von den internen Policies des Archivs variieren. Der Prüfprozess kann

feststellen, dass bestimmte Teile des SIPs für die Integration in das Archiv nicht geeignet sind und neu eingereicht oder ausgeschlossen werden müssen. Der *Prüfbericht* wird an die ‚Übernahme' abgeliefert. Nach Abschluss des Prüfprozesses werden alle *Vorbehalte* dem Produzenten mitgeteilt, der das SIP daraufhin erneut an die ‚Übernahme' schickt oder *Einspruch* gegen die Entscheidung der ‚Administration' einlegt. Nachdem die ganze Prüfung abgeschlossen ist, wird ein *abschließender Übernahmebericht* erstellt und an den Produzenten sowie an ‚Übergabevereinbarung verhandeln' geschickt. Prüfmethoden können Stichproben, periodische Prüfungen und Expertengutachten umfassen.

Die Funktion **‚Anfragen aktivieren'** führt eine Liste mit ereignisbasierten Anfragen und vergleicht diese regelmäßig mit den Inhalten des Archivs, um festzustellen, ob alle notwendigen Daten verfügbar sind. Wenn die notwendigen Daten verfügbar sind, erzeugt diese Funktion eine *Auslieferungsanfrage,* die an den ‚Zugriff' gesendet wird. Diese Funktion kann außerdem in regelmäßigen Abständen Bestellungen generieren, wobei die zeitlichen Abstände von den Endnutzern bestimmt werden können oder vom Eintreten eines Ereignisses (z.B. eine Datenbankaktualisierung) abhängig sind.

Die Funktion **‚Kunden-Dienste'** erstellt, pflegt und löscht Endnutzerkonten. Sie sammelt *Abrechnungsinformationen* aus dem ‚Zugriff', verschickt *Rechnungen* und nimmt *Zahlungen* der Endnutzer für die Benutzung von Ressourcen des Archivsystems entgegen. Sie beantwortet allgemeine Anfragen. Diese Funktion sammelt und beantwortet außerdem Feedback zu den Zugriffsdiensten und Produkten. Die ‚Kunden-Dienste' fassen diese *Kommentare* zusammen und machen sie zugänglich.

4.1.1.6. Erhaltungsplanung

Die Funktionen der Funktionseinheit ‚Erhaltungsplanung' sind in Abbildung 4-6 dargestellt.

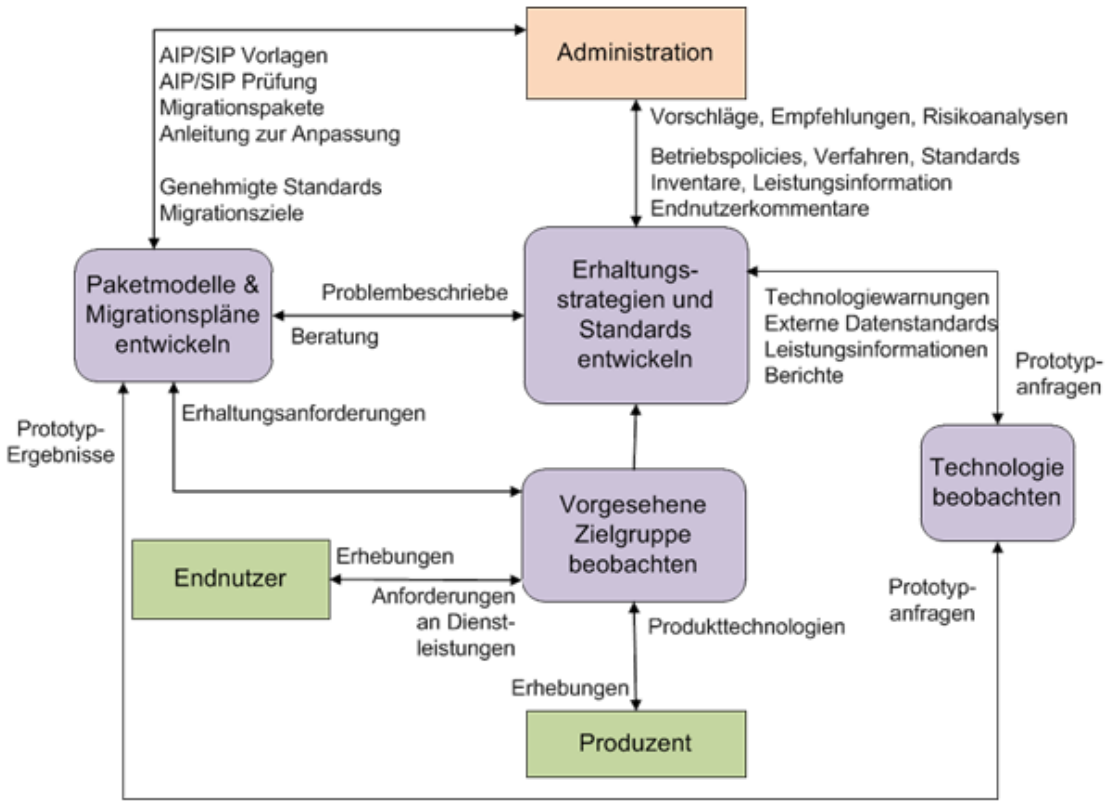

Abbildung 4-6: Funktionen der Funktionseinheit ‚Erhaltungsplanung'

Die Funktion **‚vorgesehene Zielgruppe beobachten'** steht in Kontakt mit Endnutzern und Produzenten des Archivs, um Veränderungen in deren Anforderungen an Dienstleistungen und in den verfügbaren *Produkttechnologien* zu erkennen. Solche Anforderungen können Dateiformate, die Wahl der Speichermedien, Präferenzen für Softwarepakete, neue Rechnerplattformen und Mechanismen für die Kommunikation mit dem Archiv betreffen. Diese Funktion kann mittels *Erhebungen,* über einen periodisch stattfindenden formalen Reviewprozess, in gemeinschaftlichen Workshops mit der vorgesehenen Zielgruppe, bei denen Feedback erbeten wird, oder durch individuellen Austausch durchgeführt werden. Sie liefert *Berichte, Bedarfswarnungen* und *neue Standards* an ‚Erhaltungsstrategien und Standards entwickeln'. Sie sendet *Erhaltungsanforderungen* an ‚Paketmodelle entwickeln'.

Die Funktion **‚Technologie beobachten'** ist dafür verantwortlich, die Entwicklung neu entstehender digitaler Technologien, Informationsstandards und Rechnerplattformen (d.h. Hardware und Software) zu verfolgen, um Technologien zu erkennen, welche Obsoleszenz in der Rechner-Umgebung des Archivs verursachen und die Nutzbarkeit der vorhandenen Archivbestände gefährden können. Diese Funktion kann die Fähigkeit umfassen, Prototypen zu entwickeln, um neue Technologien besser beurteilen zu können. Sie kann *Prototypanfragen* von ‚Erhaltungsstrategien und Standards entwickeln' und von ‚Paketmodelle und Migrationspläne entwickeln' erhalten. Diese Funktion sendet *Berichte, externe Datenstandards, Prototyp-Ergebnisse* und *Technologiewarnungen* an ‚Erhaltungsstrategien und Standards entwickeln'. Sie sendet außerdem *Prototyp-Ergebnisse* an ‚Paketmodell und Migrationspläne entwickeln'.

Die Funktion **‚Erhaltungsstrategien und Standards entwickeln'** ist verantwortlich für die Entwicklung und Empfehlung von Strategien und Standards und für die Bewertung von Risiken, um dem Archiv fundierte Kompromisse zu ermöglichen, wenn es Standards festlegt, Policies aufstellt und seine System-Infrastruktur verwaltet. Risiko-Management ist eine geeignete Methodologie, um einen Ausgleich zwischen Anforderungen und Mitteln und zwischen unmittelbarem Handlungsbedarf und langfristigen Zielen des Erhaltungsauftrags zu finden. Risiko-Management kann auch eine nützliche Metrik liefern, um Dinge zu quantifizieren, die in einem Entscheidungsprozess gewöhnlich schwierig einzuschätzen sind. Diese Funktion liefert periodisch *Risikoanalysen* an die ‚Administration', in denen erwartete Risiken und deren mögliche Minimierung, gestützt auf aktuelle und vorgeschlagene Aktualisierungen von Betriebsrichtlinien, Verfahren und Standards, untersucht werden. Diese Funktion beobachtet Veränderungen in den Dienstleistungsanforderungen der vorgesehenen Zielgruppe oder in Technologietrends, welche die Migration von einigen aktuellen Archivbeständen oder von neuen Übergaben erfordern können. Diese Funktion erhält *Berichte* von den Funktionen ‚vorgesehene Zielgruppe beobachten' und ‚Technologie beobachten', und sie erhält von der ‚Administration' *Betriebspolicies*, *Verfahren und Standards*, *Leistungsinformationen*, *Inventarberichte* und zusammengefasste *Endnutzer-Kommentare*. Diese Funktion sendet *Empfehlungen* zur Weiterentwicklung des Systems an die ‚Administration'. Diese Funktion erhält ebenfalls *externe Datenstandards* von ‚Technologie überwachen' und erstellt Profile von solchen Standards, die der ‚Administration' zur möglichen Anwendung vorgeschlagen werden. Diese Funktion erhält außerdem *Problembeschreibungen* von ‚Paketmodelle und Migrationspläne entwickeln' im Fall von unerwarteten Übergabeanforderungen und liefert *Beratung* bei der Bewältigung der neuen Anforderungen.

Die Funktion **‚Paketmodelle und Migrationspläne entwickeln'** erarbeitet neue Modelle für Informationspakete, detaillierte Migrationspläne und Prototypen, um die Richtlinien und Weisungen der ‚Administration' umzusetzen. Diese Aktivität gibt außerdem Beratung über die Anwendung dieser Informationspaket-Modelle und Migrationspläne auf spezifische Archivbestände und Übergaben. Die Funktion erhält von der ‚Administration' vom Archiv *geneh-*

migte Standards und *Migrationsziele*. Die Standards umfassen Formatstandards, Metadatenstandards und Dokumentationsstandards. Die Funktion wendet diese Standards auf die *Erhaltungsanforderungen* an und erstellt für die ‚Administration' *AIP- und SIP-Vorlagen*. Außerdem liefert diese Funktion im Hinblick auf die Anwendung dieser Modelle eine *Anleitung zur Anpassung* und eine *AIP/SIP-Prüfung* an die ‚Administration'. Wenn diese Funktion auf Übergaben trifft, die nicht durch die existierenden Standards und Verfahren abgedeckt sind, kann sie *Problembeschreibungen* an ‚Erhaltungsstrategien und Standards entwickeln' schicken und bekommt von ihr *Beratung*, inklusive neuen Standards, um den neuen Übergabeanforderungen zu begegnen.

Die Migrationsziele, welche diese Funktion vorgegeben bekommt, laufen auf Transformationen des AIPs hinaus, inklusive Umwandlungen der Inhaltsinformation, um zu vermeiden, dass wegen der technologischen Obsoleszenz der Zugriff unmöglich wird (siehe Abschnitt 5). Die Reaktion auf die Migrationsziele kann in der Entwicklung von neuen AIP-Modellen, von Software-Prototypen, in Testplänen, in Plänen zur Konsultation der vorgesehenen Zielgruppe und in Plänen zur stufenweisen Einführung der neuen AIPs bestehen. Dieser Prozess kann Fachkenntnisse oder Ressourcen von anderen Funktionen innerhalb der ‚Erhaltungsplanung' abrufen, z.B. die Entwicklung von Prototypen durch ‚Technologie überwachen'. Diese Aufgabe wird auch die Konsultation anderer Funktionsbereiche und der vorgesehenen Zielgruppe erforderlich machen. Sobald der Migrationsplan, das dazugehörende AIP-Modell und die Software von der ‚Administration' getestet und genehmigt sind, wird diese Funktion das gesamte *Migrationspaket* an die ‚Administration' schicken. Diese Vorschläge für den Migrationsplan werden entgegengenommen und genehmigt (oder abgelehnt) durch die Funktion ‚Standards und Richtlinien festlegen' innerhalb der ‚Administration'. Die Funktionseinheit ‚Erhaltungsplanung' entwickelt, validiert und liefert auf der Grundlage dieser Genehmigung die Migrationspakete. Die ‚Administration' erstellt einen Zeitplan und setzt die Migrationspläne um.

4.1.1.7 Zugriff

Die Funktionen der Funktionseinheit ‚Zugriff' sind in Abbildung 4-7 illustriert.

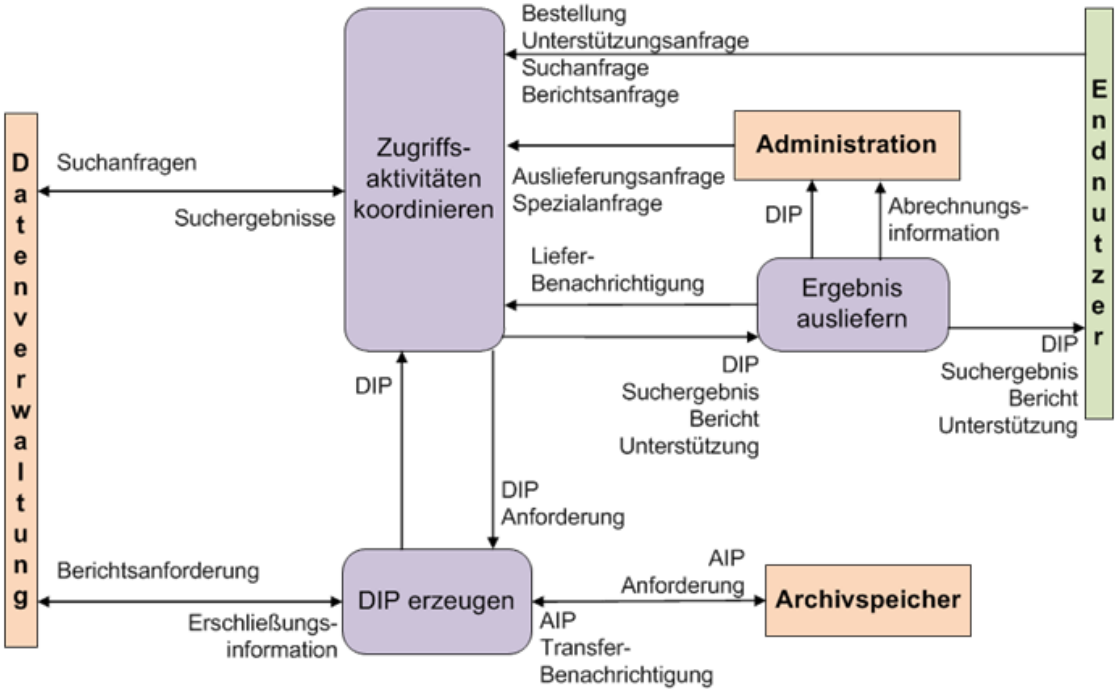

Abbildung 4-7: Funktionen der Funktionseinheit ‚Zugriff'

Die Funktion **‚Zugriffsaktivitäten koordinieren'** stellt eine oder mehrere Schnittstellen zu den Informationsbeständen des Archivs bereit. Diese Schnittstelle läuft normalerweise über ein Computernetzwerk oder eine Wählleitung zu einem Onlinedienst, sie kann aber auch in Form einer Präsenzeinrichtung, einem auf einem gedruckten Katalog basierten Bestellservice oder eines Faxabrufservices realisiert werden. Drei Kategorien von Endnutzeranfragen werden unterschieden: *Suchanfragen,* die in der ‚Datenverwaltung' ausgeführt werden und die umgehend *Suchergebnisse* zur Vorlage an den Benutzer zurückschicken; *Berichtsanforderungen,* die unter Umständen eine ganze Reihe von Anfragen nötig machen und formatierte *Berichte* zur Lieferung an den Endnutzer produzieren sowie *Bestellungen,* die entweder auf die ‚Datenverwaltung' oder den ‚Archivspeicher' oder auf beide gleichzeitig zugreifen, um ein formelles *Auslieferungsinformationspaket (DIP)* für die Online- oder Offline-Lieferung vorzubereiten. Bei einer Bestellung kann es sich um eine Spontan-Bestellung handeln, die nur ein einziges Mal ausgeführt wird, oder aber um eine ereignisbasierte Bestellung, welche von der Funktion ‚Anfragen aktivieren' innerhalb der ‚Administration' verwaltet wird und durch eine *Auslieferungsanfrage* initiiert wird, was zu regelmäßigen Lieferungen von nachgefragten Einheiten führen kann. Die Funktion ‚Archivinformation aktualisieren' in der ‚Administration' sendet ebenfalls *Auslieferungsanfragen,* um DIPs zu bekommen, die sie benötigt, damit die Aktualisierungsfunktionen ausgeführt werden kann. Andere, spezielle Anfragetypen sind erlaubt, werden aber hier nicht genauer beschrieben. Diese Funktion wird bestimmen, ob Ressourcen zur Bearbeitung einer Anfrage verfügbar sind, sicherstellen, dass der Benutzer autorisiert ist, auf die angefragten Elemente zuzugreifen und sie entgegenzunehmen, und den Endnutzer benachrichtigen, ob die Anfrage akzeptiert oder abgewiesen wurde (unter Umständen mit einem Kostenvoranschlag und der Möglichkeit, die

Anfrage zurückzuziehen). Sie wird die Anfrage anschließend zur Abwicklung an die ‚Datenverwaltung' oder an ‚DIP erzeugen' übermitteln. Darüber hinaus bietet diese Funktion auf eine entsprechende *Unterstützungsanfrage* hin den OAIS-Endnutzern *Unterstützung* und informiert sie über den Status von Bestellungen und anderen Aktivitäten des Endnutzersupports.

Die Funktion **‚DIP erstellen'** nimmt eine *Auslieferungsanfrage* entgegen, ruft das AIP im ‚Archivspeicher' ab und verschiebt eine Kopie der Daten zur weiteren Bearbeitung in einen Zwischenspeicher. Außerdem übermittelt diese Funktion eine *Berichtsanfrage* an die ‚Datenverwaltung', um die für das DIP benötigte *Erschließungsinformationen* zu erhalten. Wenn eine spezielle Verarbeitung erforderlich ist, greift die Funktion ‚DIP erstellen' auf die Datenobjekte im Zwischenspeicher zu und führt die verlangten Prozesse durch. Die Bearbeitungen, die durchgeführt werden können, umfassen statistische Funktionen, Aufteilung in den zeitlichen oder räumlichen Dimensionen, Konvertierungen zwischen verschiedenen Datentypen oder Ausgabeformaten und andere spezielle Bearbeitungen (z.B. Bildbearbeitung). Unter diese Bearbeitungstypen fallen auch die Integration von DRM-Information und die Filterung von personenbezogenen Daten, um konform mit den Nutzungsrechten zu sein. Diese Funktion stellt das fertige DIP in den Zwischenspeicherspeicherbereich und benachrichtigt die Funktion 'Zugriffsaktivitäten koordinieren', dass das DIP bereit ist zur Lieferung. Es muss darauf hingewiesen werden, dass in einigen Implementationen der AIP-Inhalt oder die DIPs in Zwischenspeichern vorgehalten werden können, um sie rasch verfügbar zu haben.

Die Funktion **‚Ergebnis ausliefern'** erledigt sowohl Online- als auch Offline-Lieferungen von Ergebnissen *(DIPs, Suchergebnisse, Berichte* und *Unterstützung)* an die Endnutzer. Für die Online-Lieferung nimmt sie ein Ergebnis von ‚Zugriffsaktivitäten koordinieren' entgegen und bereitet es für den Online-Versand in Echtzeit über Datenleitungen vor. Sie identifiziert den beabsichtigten Empfänger, bestimmt die gewünschte Übertragungsart, stellt die Antwort zur Übertragung in den Zwischenspeicherbereich und unterstützt die Online-Übertragung der Antwort. Für die Offline-Lieferung holt sie die Ergebnisse von der Funktion ‚Zugriffsaktivitäten koordinieren', erstellt eine Versandliste und andere Versandbelege und versendet anschließend die Antwort. Wenn die Antwort versendet worden ist, werden eine *Versandbestätigung* an die Funktion ‚Zugriffsaktivitäten koordinieren' zurückgeschickt und *Abrechnungsinformationen* an die ‚Administration' übermittelt.

4.1.2 Datenfluss-Diagramme

Der Fluss der Datenelemente zwischen den OAIS-Funktionseinheiten ist in diesem Unterabschnitt grafisch dargestellt. Abbildung 4-8 zeigt die zentralen Datenflüsse. Um die Abbildung nicht zu kompliziert zu machen, sind die Administrations-Datenflüsse, die in der Regel Hintergrundaktivitäten sind, in eine eigene Grafik des Administrationskontexts ausgegliedert worden: Abbildung 4-9. Datenflüsse in Verbindung mit ‚allgemeinen Diensten' sind implizit in den dargestellten Funktionen vorhanden und werden deshalb nicht angegeben.

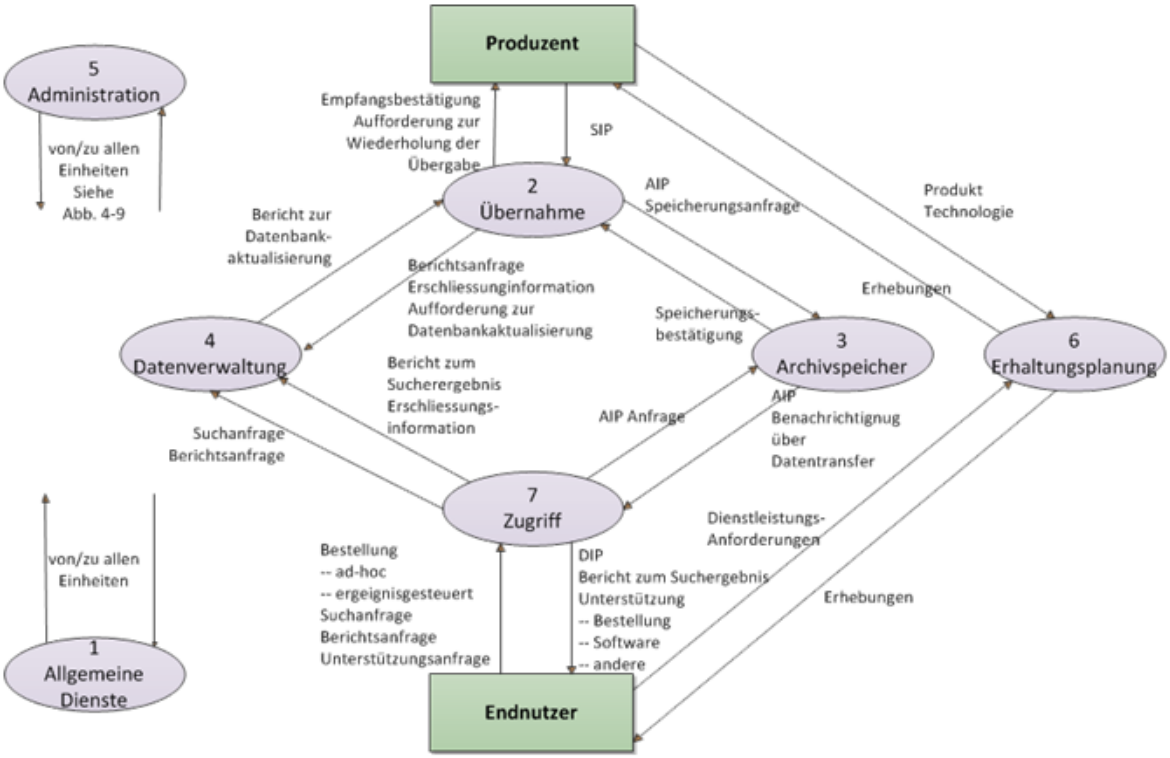

Abbildung 4-8: OAIS Datenfluss-Diagramm

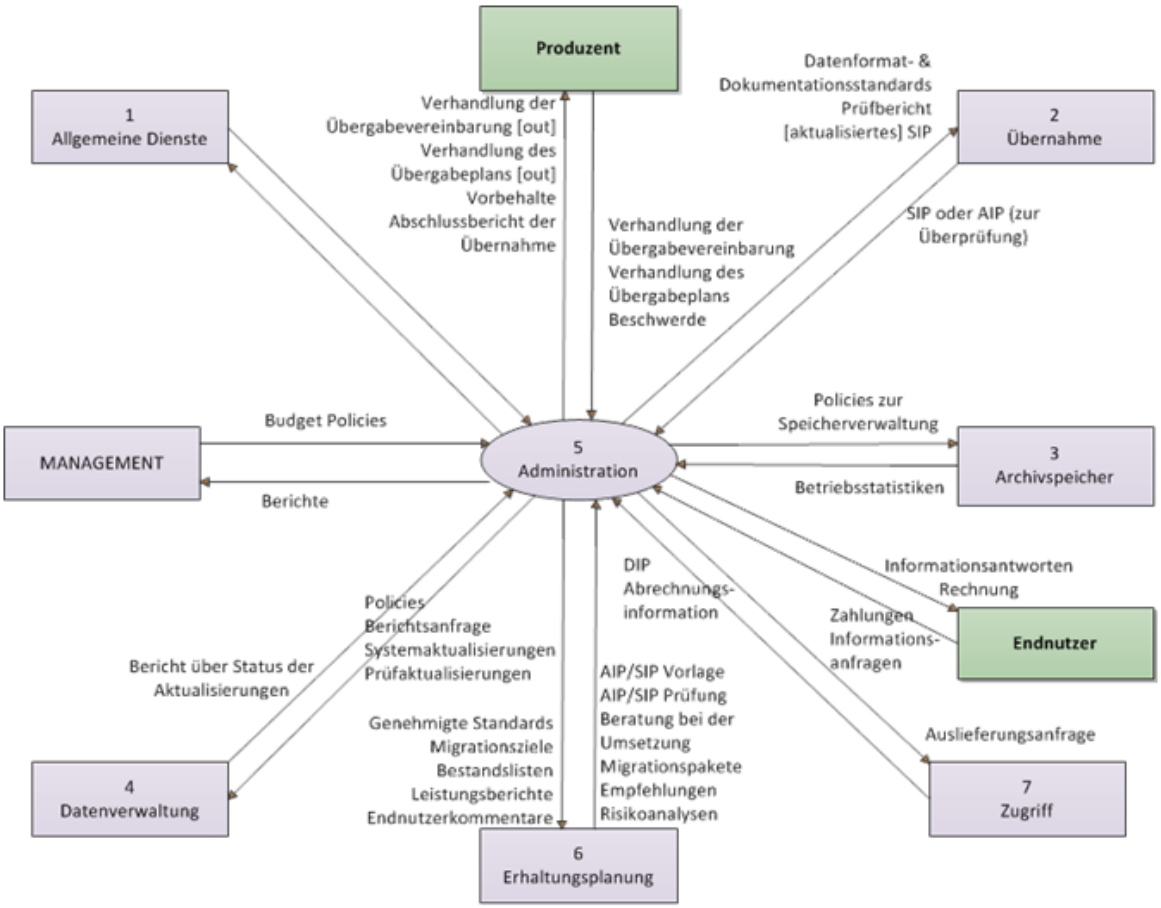

Abbildung 4-9: Diagramm des Administrationskontextes

4.2 Informationsmodell

Dieser Unterabschnitt baut auf den in Abschnitt 2 vorgestellten Konzepten auf, um die verschiedenen Arten von Information näher zu beschreiben, die im OAIS ausgetauscht und verwaltet werden. Zudem werden die spezifischen Informationsobjekte definiert, die innerhalb des OAIS genutzt werden, um die dem Archiv anvertraute Information zu erhalten und zugänglich zu machen. Dieses detailliertere Modell von OAIS-bezogenen Informationsobjekten soll Architekten oder Gestalter künftiger OAIS-Systeme helfen. Die in diesem Unterabschnitt behandelten Objekte sind rein konzeptioneller Natur und sollten nicht als Vorgabe einer konkreten Implementation verstanden werden.

Wie in Abschnitt 2 besprochen, ist es das Hauptziel eines OAIS, Information für eine vorgesehene Zielgruppe über einen unbestimmten Zeitraum hinweg zu bewahren. Um diese Information zu erhalten, muss ein OAIS wesentlich mehr als nur die Inhalte des zu archivierenden Objekts speichern. Dieser Unterabschnitt analysiert die Informationsanforderungen, die zur Beschreibung der Objektklassen von Daten gebraucht werden, die zu einem OAIS gehören. Dieser Unterabschnitt benutzt Objektmodelldiagramme in Unified Modeling Language (UML), um die im Text diskutierten Konzepte zu illustrieren. Anhang B gibt einen Überblick über die benutzte Notation und grundlegende Konzepte der Objektmodellierung. Ein Verständnis dieser Notation ist notwendig, um die in diesem Unterabschnitt vorgestellten Konzepte in Gänze zu verstehen.

Unterabschnitt 4.2.1 stellt ein Modell der Information vor, die für eine effektive Langzeiterhaltung von Information erforderlich sind. Unterabschnitt 4.2.2 beschreibt die konzeptionellen Objekte und Container, die den Inhalt eines OAIS repräsentieren.

4.2.1. Logisches Modell der Archivinformationen

4.2.1.1 Informationsobjekt

Ein grundlegendes Konzept des OAIS-Referenzmodells ist das Konzept von Information als einer Kombination aus Daten und Repräsentationsinformation. Das UML-Diagramm in Abbildung 4-10 illustriert dieses Konzept. Das Informationsobjekt besteht aus einem Datenobjekt, das entweder physisch oder digital ist, und der Repräsentationsinformation, welche erst ein vollständiges Verständnis der Daten als bedeutungstragende Information ermöglicht. Dieses Modell ist für alle Arten von Information in einem OAIS gültig.

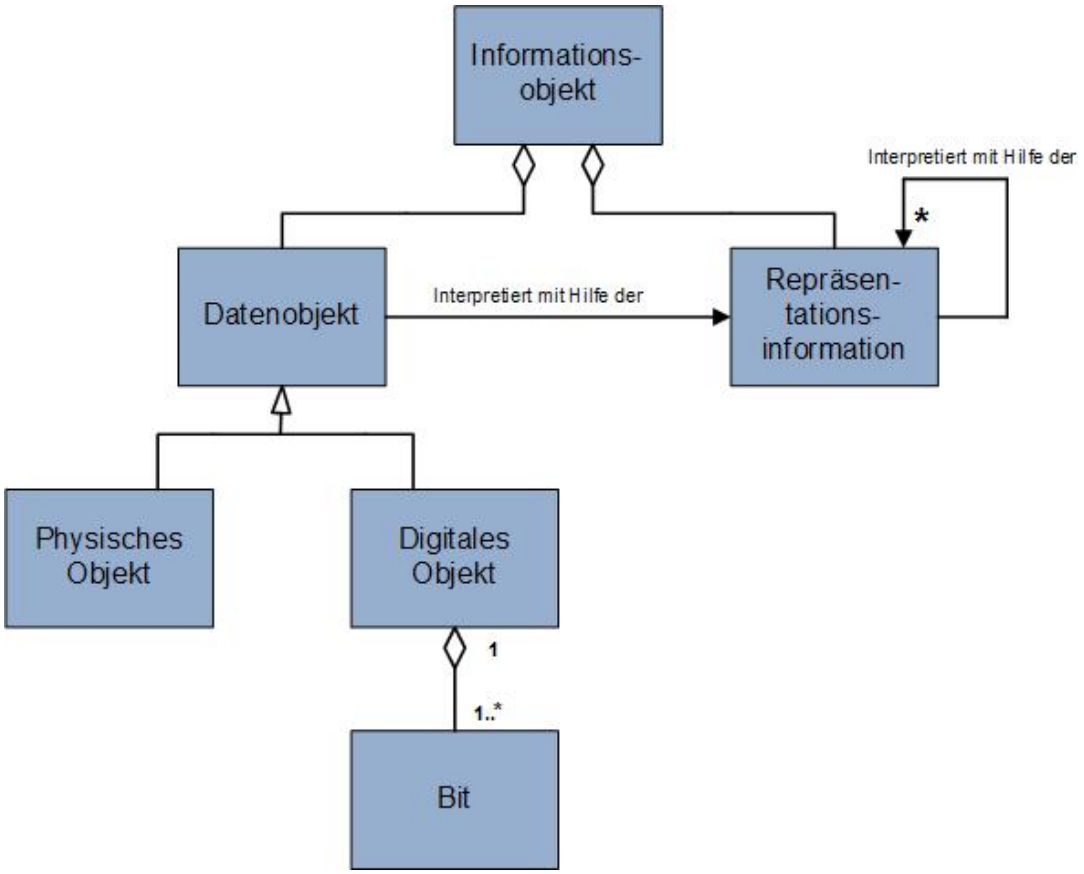

Abbildung 4-10: Informationsobjekt

4.2.1.2 Datenobjekt

Das Datenobjekt kann entweder als physisches Objekt (z.B. ein Mondstein) zusammen mit einigen Repräsentationsinformationen vorliegen oder als ein digitales Objekt (z.B. eine Bitsequenz) zusammen mit den Repräsentationsinformationen, welche diesen Bits Bedeutung geben.

4.2.1.3. Repräsentationsinformation

Die Repräsentationsinformation, die ein digitales Objekt (d.h. eine Bitsequenz) begleitet, dient dazu, zusätzliche Bedeutung hinzuzufügen. Üblicherweise bildet sie die Bits auf allgemein bekannte Datentypen ab, wie etwa Buchstaben, ganze Zahlen, Dezimalzahlen, und Gruppen von diesen Datentypen. Sie verknüpft diese mit Bedeutung auf höherer Ebene: Dies schließt die Beschreibung der möglicherweise komplexen Art und Weise ein, wie die Sachen zusammenhängen. (Repräsentationsinformation kann beispielsweise angeben, dass drei Zahlen Temperatur, Breitengrad und Längengrad darstellen und als Grad Celsius und Grad eines Winkels ausgedrückt werden; und sie hängen insofern zusammen, dass die Temperatur am angegebenen Breiten- und Längengrad gemessen wird.)

Die zu einem physischen Objekt – wie etwa einem Mondstein – beigegebenen Repräsentationsinformationen können als Untersuchungsresultat den beobachtbaren physikalischen Eigenschaften zusätzliche Bedeutung geben. Diese Information kann über die Zeit weiterentwickelt worden sein, und die Resultate werden, falls mitgegeben, Teil des Informationsobjekts.

Der Rest dieses Unterabschnitts konzentriert sich auf Repräsentationsinformationsobjekte, deren Datenobjekte auf digitale Objekte eingeschränkt sind.

4.2.1.3.1. Typen von Repräsentationsinformation

Wie in Abbildung 4-10 gezeigt wird, ist das digitale Objekt aus einer oder mehreren Bitfolgen zusammengesetzt. Der Zweck des Repräsentationsinformationsobjekts ist es, die Bitfolgen in aussagekräftigere Information zu übersetzen. Es leistet dies durch die Beschreibung des Formats beziehungsweise der Datenstrukturkonzepte, die auf diese Bitfolgen angewendet werden, womit diese zu aussagekräftigeren Werten wie Zeichen, Zahlen, Pixel, Felder, Tabellen, etc. werden. Diese geläufigen Computerdatentypen, Gruppierungen dieser Datentypen sowie die Abbildungsregeln, welche die zugrundeliegenden Datentypen auf höherstufige Konzepte abbilden, die für ein Verständnis des digitalen Objekts notwendig sind, werden als **Strukturinformation** des Repräsentationsinformationsobjekts bezeichnet. Diese Strukturen werden im Allgemeinen durch Namen identifiziert oder durch ihre relative Position innerhalb der Bitsequenz. Die Strukturinformation wird häufig als das "Format" des digitalen Objekts bezeichnet.

Die Repräsentationsinformationen, die von den Strukturinformationen geliefert werden, sind selten ausreichend. Sogar wenn ein digitales Objekt aus einer Buchstabenfolge besteht und in der Strukturinformation als solche beschrieben wird, sollte die Zusatzinformation bereitgestellt werden, in welcher Sprache der Text geschrieben ist. Diese Art von benötigten Zusatzinformationen wird **semantische Information** genannt. Bei Wissenschaftsdaten zum Beispiel kann die Information in der semantischen Information sehr vielseitig und komplex sein. Sie beinhaltet die speziellen Bedeutungen, die mit allen Elementen der Strukturinformation verknüpft sind, die auf jedem Datentyp durchführbaren Operationen sowie ihre Beziehungen untereinander. Abbildung 4-11 unterstreicht die Tatsache, dass die Repräsentationsinformationen sowohl Strukturinformation als auch semantische Information enthalten, auch wenn der Unterschied in einigen Umsetzungen subjektiv ist. Es ist hilfreich, sich zu vergegenwärtigen, dass die semantische Information, die mit bestimmter digital codierter Information verknüpft ist, üblicherweise unabhängig vom Format ist. Beispielsweise ist die Bedeutung von Zahlen in einer Datei unabhängig davon, ob sie als Festkommazahlen oder als Gleitkommazahlen kodiert sind; die Bedeutung von Wörtern in einem Dokument ist unabhängig davon, ob das Dokument in Word oder PDF gespeichert ist.

Diese Abbildung zeigt ferner, dass Repräsentationsformation auch sonstige Repräsentationsinformation enthalten kann. Dies deutet darauf hin, dass die hier dargestellte Taxonomie von Repräsentationsinformation alles andere als vollständig ist und unter Umständen z.B. Software, Algorithmen, Verschlüsselung, Bedienungsableitungen und viele andere Dinge zum Verständnis des Inhaltsdatenobjekts benötigt werden, die deshalb alle nach Definition Repräsentationsinformation, aber nicht eindeutig Strukturinformation oder semantische Information wären. Z.B. würde Information, die definiert, wie Strukturinformation und semantische Information sich zueinander verhalten, oder zur Verarbeitung einer Datenbankdatei benötigte Software als sonstige Repräsentationsinformation betrachtet werden. Strukturinformation, semantische Information und sonstige Repräsentationsinformation sind sowohl Bestandteile als auch Untertypen der Repräsentationsinformation.

Die Repräsentationsinformation ist ihrerseits ein Informationsobjekt, das sein eigenes digitales Objekt und weitere Repräsentationsinformationen haben kann, wie sie mit dem Verständnis jedes digitalen Objekts verknüpft ist. Dies ist in kompakter Form durch die „interpretiert mit Hilfe der“-Beziehung dargestellt. Das daraus entstehende Geflecht von Objekten kann als **Repräsentationsnetzwerk** bezeichnet werden.

Zum Beispiel beschreibt die ISO-Norm 9660 [11], dass Text dem ASCII-Standard entspricht, aber sie beschreibt nicht, wie ASCII umgesetzt werden muss. Sie verweist nur auf den ASCII-Standard, der somit eine zusätzliche Repräsentationsinformation ist, die für ein volles Verständnis gebraucht wird. Deshalb ist der ASCII-Standard Teil des mit ISO 9660 verbundenen Repräsentationsnetzes. Das OAIS muss ihn sich entweder in irgendeiner Form verschaffen oder seine Verfügbarkeit im Auge behalten, so dass es in Zukunft die angemesse-

nen Maßnahmen ergreifen kann, um sicherzustellen, dass seine Repräsentationsinformationen zu ISO 9660 auch vollständig verstehbar sind.

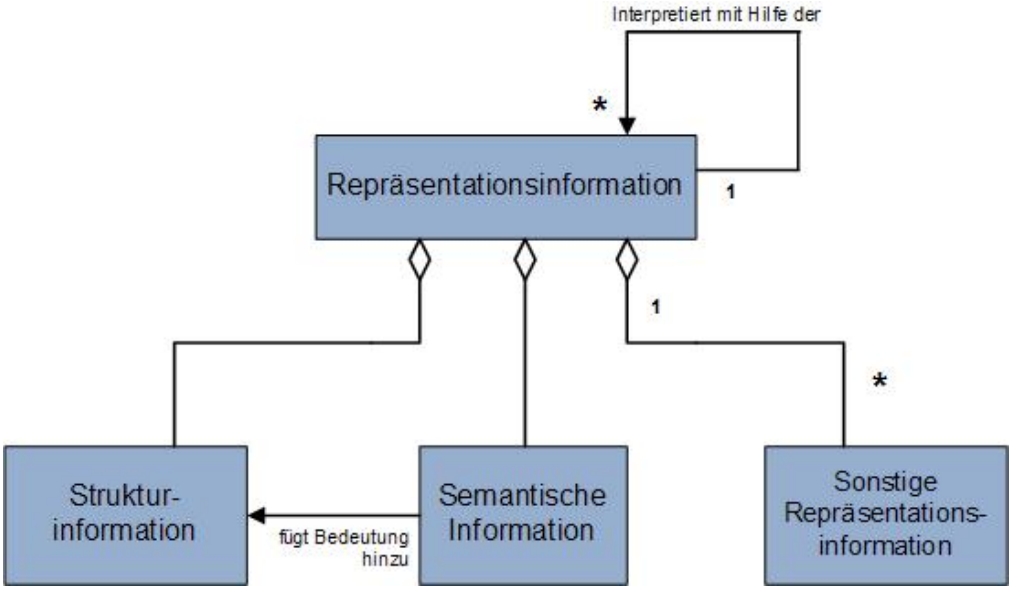

Abbildung 4-11: Repräsentationsinformationsobjekt

4.2.1.3.2 Repräsentationsnetzwerke

Repräsentationsinformation, die ja selbst ein Informationsobjekt ist, kann in analoger Form existieren (z.B. als Papierdokument) oder in digitaler Form. Wenn die Repräsentationsinformation in digitaler Form vorhanden ist, wird zusätzliche Repräsentationsinformation benötigt, um die Bits der Repräsentationsinformation verstehen zu können, wie im vorherigen Unterabschnitt beschrieben. Prinzipiell hält dieser Rückbezug solange an, bis analoge Formen auftreten, die von der vorgesehenen Zielgruppe verstanden werden können. Zum Beispiel: Die Repräsentationsinformationen, die in ASCII ausgedrückt werden, brauchen zusätzliche Repräsentationsinformationen für ASCII. Das kann ein analoges Dokument sein, das den ASCII-Standard beschreibt. Jede Repräsentationsinformation kann mehrere Komponenten haben, darunter auch mehrere referenzierte Komponenten Repräsentationsinformation; jede davon wieder mit eigenen Repräsentationsinformationen.

Um die Bedeutung eines Informationsobjekts zu erhalten, muss auch seine Repräsentationsinformation erhalten bleiben. Das wird am leichtesten erreicht, indem die Repräsentationsinformationsobjekte in leicht verstehbaren Formen ausgedrückt werden, zum Beispiel als Textbeschreibungen, die weit verbreitete Standards wie die ASCII-Zeichen für elektronische Versionen verwenden. Ein Problem bei der Verwendung von reinen Textbeschreibungen ist, dass solche Beschreibungen doppeldeutig sein können. Eine Art des Umgangs damit ist, standardisierte, formelle Beschreibungssprachen zu verwenden, die wohl definierte Konstrukte enthalten, mit deren Hilfe Datenstrukturen beschrieben werden. Diese Sprachen müssen eventuell angereichert werden mit Textbeschreibungen, um die Semantik der Repräsentationsinformationen voll zu übermitteln.

Da sich das Grundwissen der vorgesehenen Zielgruppe mit der Zeit ändert, kann sich das Repräsentationsnetzwerk auch entsprechend ändern müssen. Wie in 2.2 bemerkt, hat das OAIS die Wahl, alle relevante Repräsentationsinformation selber zu sammeln oder auf deren Existenz in einem anderen vertrauenswürdigen oder Partner-OAIS-Archiv zu verweisen; dies ist eine Umsetzungs- und Organisationsentscheidung.

Die Inhaltsinformation muss definiert werden und in Inhaltsdatenobjekt und Repräsentationsinformation getrennt werden. Auch dies ist eine Umsetzungs- und Organisationsentscheidung, die damit zusammenhängt, wie Datenobjekte ins OAIS übernommen und dort gespeichert werden. Beispielsweise könnte im Fall der darstellenden Künste das Inhaltsdatenobjekt eine Partitur als PDF-Dokument sein, und die Repräsentationsinformation würde dann alle Informationen enthalten, die benötigt würden, um das Stück wieder aufzuführen (als eine Art, es zu nutzen und zu verstehen). Dazu könnten die Art der Darstellung der PDF-Datei gehören, die benötigte Audioverarbeitungssoftware, die Positionierung der Hardware (Lautsprecher u.a.), Bewegungsanweisungen und eine Beschreibung, wie sich diese zueinander und zum Inhaltsdatenobjekt verhalten, wobei all dies sehr komplex sein, auf eigene Art kodiert und nicht einfach als entweder Struktur oder Semantik beschrieben werden kann. Alternativ kann das Inhaltsdatenobjekt mehrere Datenobjekte umfassen, die die Partitur, die benötigte Audioverarbeitungssoftware, Hardwarepositionierung und Bewegungsanweisungen beinhalten. Jedes dieser Datenobjekts braucht seine eigene Repräsentationsinformation, und es wird zusätzliche Repräsentationsinformation benötigt, die beschreibt, wie die verschiedenen Datenobjekte sich zueinander verhalten.

Zwei besondere Arten der Repräsentationsinformation sind das **Repräsentations-Wiedergabe-Programm** und das **Zugriffsprogramm**. Das Repräsentations-Wiedergabe**-Programm** kann die Repräsentationsinformationen in verstehbarer Form darstellen. Ein Beispiel: Die Datei- und Verzeichnisstruktur vieler CD-ROMs entspricht der ISO-Norm 9660. Dieser Standard ist Repräsentationsinformation, die beschreibt, wie die meisten Dateistrukturen auf CD-ROMs umgesetzt werden müssen, und ist als Papierdokument erhältlich. Der Standard kann jedoch auch als digitales Objekt, als PDF-Datei, bezogen werden. Statt die Dokumentation von PDF zu archivieren und eine Software zu schreiben, um das digitale Objekt "ISO 9660" zu verstehen, kann ein OAIS eine vorhandene PDF-Anzeige-Software verwenden, um die ISO 9660-Dokumentation für Menschen sichtbar und lesbar darzustellen. In diesem Fall wird die Software, die das PDF öffnet, als Repräsentations-Wiedergabe-Programm bezeichnet, weil sie zur Darstellung der Repräsentationsinformationen benutzt wird. Wenn das OAIS nicht auch die Begleitbeschreibung von PDF archiviert, muss es diese Tatsache dokumentieren und im Blick behalten, denn sobald es kein kosteneffektive Möglichkeit mehr gibt, PDF-Dateien zu öffnen und anzuzeigen, muss die als PDF-Objekt ausgedrückte ISO 9660-Dokumentation in eine neue Form migriert werden.

Zugriffsprogramme stellen einige oder alle Informationen eines Informationsobjekts in einer Form dar, die für Menschen oder Systeme verstehbar ist. Sie können auch bestimmte Zugriffsdienste zur Verfügung stellen, wie z.B. die Anzeige, das Verändern, das Verarbeiten oder Unterteilen von Informationsobjekten. Für einige Arten von digitalen Objekten können solche Programme weit verbreitet sein, und es ist deshalb nicht notwendig, dass das OAIS sie selber pflegt oder bereitstellt. Für speziellere Arten von digitalen Objekten hingegen kann es im Interesse des OAIS liegen, solche Software zu pflegen und bereitzustellen.

Weil das Zugriffsprogramm die Repräsentationsinformationen in gewissem Umfang versteht, könnten einige Archive versucht sein, das Zugriffsprogramm als Ersatz für vollständige Repräsentationsinformationen einzusetzen. Der Quellcode eines Zugriffsprogramms, der zumindest ein teilweises Verständnis der zugehörigen Repräsentationsinformationen verkörpert, könnte als Dokumentation verwendet werden, die solche Repräsentationsinformationen ausdrückt. Problematisch bei dieser Herangehensweise ist, dass die gewünschten Repräsentationsinformationen möglicherweise nicht klar identifizierbar sind, weil sie vermischt sind mit verschiedenen Algorithmen für die Verarbeitung und die Darstellung, und dass sie unvollständig sein könnten, weil der Quellcode eine gewisse Betriebsumgebung voraussetzt. Es kann schwierig sein, anhand des Softwarecodes zu bestimmen, welche Repräsentationsinformationen fehlen. Die Verwendung eines funktionsfähigen Zugriffsprogramms ohne Quellcode, wie es bei proprietären Formaten der Fall sein kann, stellt ein wesentlich größeres Risiko für Informationsverlust dar, da es schwieriger ist, eine funktionierende Umgebung für die Software aufrechtzuerhalten, als die Dokumentation über die Zeit zu migrieren. Der

Einsatz von Emulationstechniken zur Erhaltung von funktionierender Software ist ein aktiver Forschungszweig. Das ist ein gewichtiges Thema für alle, denen der Erhalt des Look And Feel des Informationszugriffs wichtig ist. Migration und die Erhaltung von Software werden ausführlicher in Abschnitt 5 diskutiert.

4.2.1.4 Die Taxonomie der Informationsobjektklassen im OAIS

Es gibt viele Informationstypen, die bei der Langzeiterhaltung von Information in einem OAIS eine Rolle spielen. Jede dieser Arten kann als komplettes Informationsobjekt betrachtet werden, denn sie enthält ein Datenobjekt und die entsprechenden Repräsentationsinformationen zum Verständnis der Daten. Dieser Unterabschnitt baut einerseits auf der Diskussion in 2.2 über die Arten von unterstützender Information auf, die gebraucht werden, um die Langzeiterhaltung zu ermöglichen, anderseits auf der Diskussion in dem vorherigen Unterabschnitt über die Rolle der Repräsentationsinformationen. Die Informationsmodellierung in diesem Unterabschnitt behandelt verschiedene Arten von Informationsobjekten, die in einem OAIS benutzt werden. Die Objekte werden nach ihrem Inhalt und ihrer Funktion im Betrieb eines OAIS eingeteilt in Inhaltsinformationsobjekte, Erhaltungsmetadatenobjekte, Verpackungsinformationsobjekte und Erschließungsinformationsobjekte. Die folgenden Unterabschnitte behandeln den Inhalt aller dieser Arten von Informationsobjekten. Abbildung 4-12 zeigt eine Systematik der im OAIS benutzten Informationsobjekte.

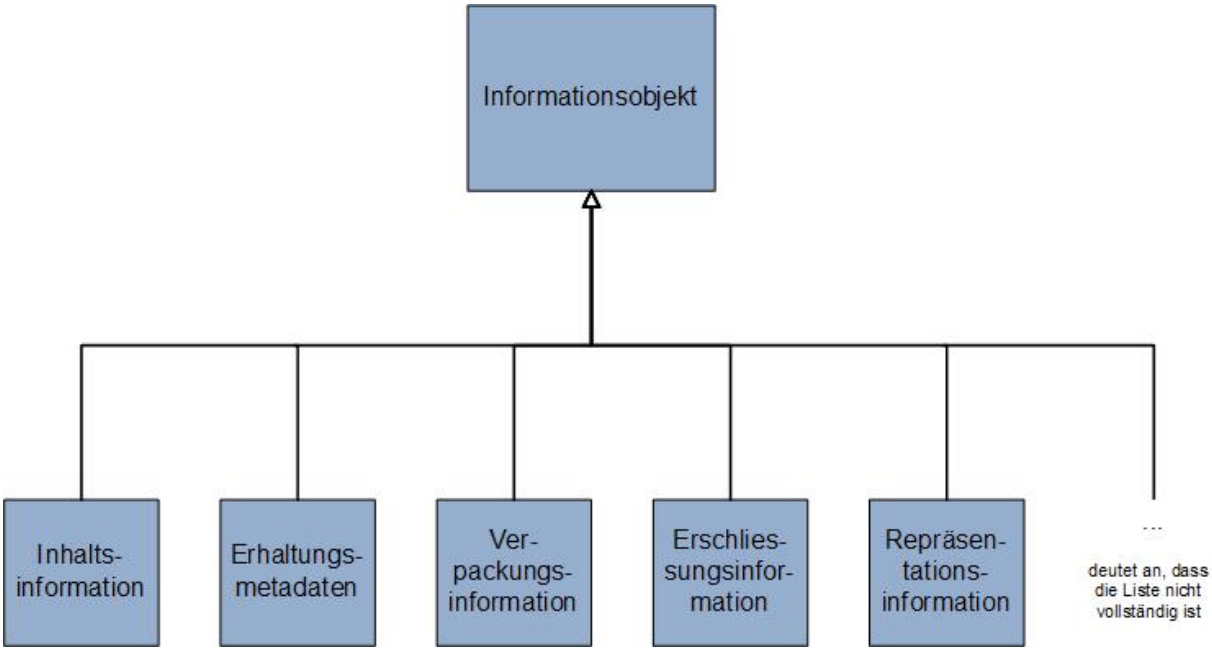

Abbildung 4-12: Taxonomie der Informationsobjekte

4.2.1.4.1 Inhaltsinformation

Die **Inhaltsinformation** ist die Information, die das eigentliche Ziel der Erhaltung durch ein OAIS ist. Die Entscheidung darüber, was genau die Inhaltsinformation ist, ist nicht immer offensichtlich und muss möglicherweise mit dem Produzenten verhandelt werden. Die Inhaltsinformation ist, wie in Abbildung 4-12 gezeigt, ein Informationsobjekt und besteht deshalb aus dem Inhaltsdatenobjekt zusammen mit den zugehörigen Repräsentationsinformationen. Das Inhaltsdatenobjekt in der Inhaltsinformation kann entweder ein digitales Objekt oder ein physisches Objekt (z.B. eine Probe oder ein Mikrofilm) sein. Jedes Informationsobjekt kann als Inhaltsinformation dienen.

Die (semantischen und syntaktischen) Repräsentationsinformationen zu einem digitalen Inhaltsdatenobjekt werden gebraucht, um die Bits vollständig in Inhaltsinformation umzuwandeln. Im Prinzip schließt das sogar die Definitionen (z.B. Wörterbuch und Grammatik) jeder natürlichen Sprache (z.B. Englisch) ein, in der die Inhaltsinformation ausgedrückt wird. Über längere Zeiträume können sich die Bedeutungen von Bezeichnungen in natürlichen Sprachen signifikant verändern und zwar sowohl im allgemeinen Sprachgebrauch als auch in spezifischen Fachsprachen.

In der Praxis braucht ein OAIS genügend Repräsentationsinformationen zu den Bits der Inhaltsdatenobjekte in der Inhaltsinformation, sodass es sich sicher ist, dass die Mitglieder der vorgesehenen Zielgruppe mit ausreichend Wissen auf das Repräsentationsnetzwerk zugreifen können, um die Repräsentationsinformationen korrekt interpretieren zu können. Dies ist ein gewichtiger Risikofaktor für ein OAIS, vor allem, wenn es Experten als vorgesehene Zielgruppe hat, weil die Fachsprache und vermeintlich weit verbreiteten Begriffe kurzlebig sein können. In solchen Fällen muss besonders sorgfältig darauf geachtet werden, dass durch die Fortentwicklung des Grundwissens der vorgesehenen Zielgruppe kein Informationsverlust bei der Inhaltsinformation auftritt.

Wie oben für ein Informationsobjekt im Allgemeinen beschrieben, kann auch ein Zugriffsprogramm, das die Präsentation der Inhaltsinformation für den Endnutzer unterstützt, als Erweiterung der Repräsentationsinformation aufgefasst werden. Beispiele für diese Art von Software sind Textverarbeitungsprogramme, die es ermöglichen, die Inhaltsinformation als komplexe Dokumentenformate darzustellen oder wissenschaftliche Visualisierungssysteme, die Inhaltsinformationen als Zeitreihe oder mehrdimensionale Datenreihe darstellen. Ein Zugriffsprogramm kann Werkzeuge zur Rechteverwaltung enthalten, die den Zugriff auf geschützte Inhalte erlauben. Die Software verwendet ihr Wissen über die zugrundeliegenden Repräsentationsinformationen, um diese Dienste bereitzustellen.

Oft ist notwendige Information in die Softwarepakete eingebettet, die die vorgesehene Zielgruppe verwendet, um die Inhaltsinformation darzustellen und zu analysieren. Ein Grund für den Erhalt funktionierender Zugriffsprogramme ist die Benutzerfreundlichkeit. Sogar mit einem kompletten Satz von Repräsentationsinformationen benötigt man für den tatsächlichen Zugriff auf alle oder einen Teil der digitalen Inhaltsdatenobjekte ein Zugriffsprogramm. Deshalb kann ein Softwaremodul, das einen nützlichen Zugriff auf ein digitales Inhaltsdatenobjekt erlaubt, aus Gründen der Benutzerfreundlichkeit lauffähig erhalten werden.

Das ist nicht schwierig, solange die Umgebung, die dieses Softwaremodul unter-stützt, einfach verfügbar ist. Diese Umgebung besteht aus zugrundeliegender Hardware und einem Betriebssystem, einigen Dienstprogrammen, die das Betriebssystem gewissermaßen erweitern, sowie aus Speicher und Anzeigegeräten und ihrer Treibersoftware. Eine Veränderung bei einer dieser Komponenten kann dazu führen, dass das Softwaremodul nicht mehr funktioniert, nicht richtig funktioniert oder der Anwendung oder dem Benutzer keine Ergebnisse mehr liefern kann. Die Komplexität dieser Interaktionen macht üblicherweise die Erhaltung einer funktionierenden Software zu einer so anstrengenden Aufgabe.

Zusammenfassend kann man sagen, dass die Verwendung von Zugriffsprogrammen als Ersatz für Repräsentationsnetzwerke attraktiv ist, um den Aufwand für die Datenübernahme und die Zugangsgewährung für die aktuellen Benutzer zu reduzieren. Aber die Abhängigkeit von funktionierender Software kann gewichtige Probleme für die Langzeiterhaltung darstellen, wenn diese Software aufhört zu funktionieren. Die unbegrenzte Langzeiterhaltung von Information erfordert eine volle und verstehbare Beschreibung der Repräsentationsinformationen. Unterabschnitt 5.2. ('Erhaltung des Zugriffs- und Nutzungsservice') behandelt einige Techniken, die angewendet werden können, um Software über die Zeit zu erhalten, sowie die mit dem Ansatz verbundenen Risiken.

Eine wichtige Funktion eines OAIS ist es, zu entscheiden, welche Teile der Inhaltsinformation das Inhaltsdatenobjekt bilden und welche Teile die Repräsentationsinformationen. Dieser Aspekt ist wesentlich für ein klares Verständnis davon, was archiviert wird. Die Bestimmung digitaler Inhaltsinformation und ihrer Repräsentationsinformationsobjekte kann durch eine Reihe von Schritten geschehen:

1) Ermitteln Sie die Bits, aus denen das Inhaltsdatenobjekt der Inhaltsinformation besteht.

2) Ermitteln Sie ein Repräsentationsinformationsobjekt, das alle Bits des Inhaltsdatenobjekts auf irgendeine Weise abdeckt und diese Bits in eine aussagekräftigere Information umwandelt.

3) Ist das Repräsentationsinformationsobjekt ermittelt, muss sein Inhalt untersucht werden, um herauszufinden, ob es weitere Repräsentationsinformationsobjekte braucht. Falls ja, besorgen Sie die notwendigen Repräsentationsinformationsobjekte. Wiederholen Sie diesen Schritt, bis keine zusätzlichen Repräsentationsinformationsobjekte benötigt werden.

4) Für jedes Repräsentationsinformationsobjekt, von dem in Punkt 3 die Rede ist und das als digitales Objekt vorgehalten wird, ermitteln Sie alle notwendigen Repräsentationsinformationsobjekt und wiederholen Sie Schritt 3 und 4, bis kein neues Repräsentationsinformationsobjekt mehr bestimmt werden kann.

5) Die Inhaltsinformation besteht aus dem Inhaltsdatenobjekt und allen Repräsentationsinformationsobjekten, die in den Schritten 2 bis 4 ermittelt wurden.

Als Beispiel für dieses Verfahren stelle man sich eine Datei vor, die eine Folge von Werten enthält, die von einem Sensor stammen, der die Umwelt beobachtet. Außerdem liegt eine zweite, in ASCII kodierte Datei vor, die Informationen darüber liefert, wie die erste Datei zu verstehen ist. Sie beschreibt, wie man die Bits der ersten Datei interpretiert, um aussagefähige Zahlen zu erhalten. Sie erklärt in Begriffen aus der Physik der durchgeführten Beobachtung, was diese Zahlen bedeuten. Sie liefert das Datum und den Zeitraum, in dem die Beobachtungen gemacht wurden, einen Durchschnittswert der beobachteten Werte und wer die Beobachtungen gemacht hat. Diese beiden Dateien werden dem OAIS zur Erhaltung übermittelt.

Angenommen, das OAIS beschließt, dass die zu archivierende Inhaltsinformation sich zusammensetzt aus den beobachteten Bits sowie ihren Werten als Zahlen und der physikalischen Bedeutung dieser Zahlen. Diese Information wird ausgedrückt durch die Bitsequenz innerhalb der ersten Datei kombiniert mit denjenigen Repräsentationsinformationen der zweiten Datei, die gebraucht werden, um die Bits der ersten Datei in aussagekräftige physikalische Werte zu verwandeln. Weder der Datenträger der ersten Datei noch das Dateisystem, in dem die Bits gespeichert sind, ist in diesem Beispiel Teil der Inhaltsinformation. Nur ein Teil des Inhalts der zweiten Datei wird als Teil der Inhaltsinformation betrachtet, und zwar der Teil, der die Transformation der Bits der ersten Datei in aussagekräftige physikalische Werte ermöglicht. Genau genommen beinhaltet diese zweite Datei nicht alle für die Umwandlung benötigten Repräsentationsinformationen, weil folgende zusätzliche Information gebraucht werden:

- die Information, dass die zweite Datei in ASCII kodiert ist, damit man sie als aussagekräftige Zeichen lesen kann;

- Information darüber, wie die Schriftzeichen verwendet werden, um die Umwandlung von Bits in Zahlen und dann in aussagekräftige physikalische Werte auszudrücken.

Diese Information, die typischerweise als eine Kombination von Formatinformation und Datenlexikoninformation bezeichnet wird, kann auch die Werte der Instrumentenkalibrierung beinhalten sowie Information darüber, wie diese Kalibrierung vorgenommen wird. Alle diese Information kann einfach verstehbar sein, sobald die ASCII-Schriftzeichen sichtbar sind, weil alles in Englisch ausgedrückt ist (oder in einer anderen natürlichen Sprache), oder aber Teile davon können in einer stärker strukturierten Form vorliegen, deren Verständnis zusätzliche Repräsentationsinformationen erfordert.

Deshalb benötigen die Repräsentationsinformationen der zweiten Datei zusätzliche Repräsentationsinformationen, die wiederum weitere Repräsentationsinformationen benötigen können usw., sodass ein vernetzter Satz von Repräsentationen von Repräsentationen entsteht. Das ist ein gutes Beispiel für ein komplexes Repräsentationsnetz.

In dem oben genannten Beispiel wird angenommen, dass die Inhaltsinformation aus den Beobachtungswerten und ihrer Bedeutung bestehen. Aber das ist keineswegs die einzige mögliche Festlegung. Man könnte genauso gut beschließen, dass das Inhaltsdatenobjekt der gewünschten Inhaltsinformation aus der Bitsequenz innerhalb der ersten Datei zusammen mit allen Bitsequenzen innerhalb der zweiten Datei besteht. Die Tatsache, dass einige von letzteren benutzt werden, um die Bitsequenzen der ersten Datei zu interpretieren, ist nur ein Beispiel für einen Satz von Bits, der zu einem gewissen Grad selbstbeschreibend ist. Es spielt keine Rolle, dass einige der Bits in der zweiten Datei die Basis für Informationen über das Datum und den Zeitraum der Messungen sind, über den Durchschnittswert der beobachteten Werte und darüber, wer die Beobachtungen gemacht hat. Sobald festgestellt wurde, dass alle diese Bits das Inhaltsdatenobjekt der Inhaltsinformation bilden, dann sind die Repräsentationsinformationen diejenige Information, die gebraucht wird, um sie in aussagekräftige Information umzuwandeln. Wie ausführlich dieses Verständnis hergestellt werden muss und wie umfangreich die Repräsentationsnetzwerke sein sollen, entscheidet jedes OAIS jeweils individuell gemeinsam mit seinen Produzenten und Endnutzern.

Als ein anderes Beispiel stelle man sich eine Datei vor, die ein Textverarbeitungsdokument enthält. Das binäre Datenobjekt hat ein komplexes Format, das nur dann als Dokument erkannt werden kann, wenn es mit Hilfe der dazugehörigen Repräsentationsinformationen betrachtet wird. Normalerweise wird diese Betrachtung von einem Zugriffsprogramm ermöglicht, das interne oder externe Repräsentationsinformationen benutzen kann. Das Inhaltsdatenobjekt wird wahrscheinlich definiert als die Bitfolge der Datei. Die Repräsentationsinformationen umfassen mindestens eine Beschreibung des Textverarbeitungsformates und allenfalls weitere Information, die als notwendig erachtet wird, um die Bedeutung des Dokuments angemessen zu verstehen. Wenn das Textverarbeitungsformat proprietär ist und nicht ausreichend Repräsentationsinformationen besorgt werden können, um zumindest die einfache Darstellung zu ermöglichen, dann kann es notwendig sein, das Dokument in ein anderes (vielleicht nicht-proprietäres) Format zu migrieren, für das Repräsentationsinformationen einfacher erhältlich sind, um seine Langzeiterhaltung zu gewährleisten.

Als Variation des obigen Beispiels kann entschieden werden, dass nicht die gesamte Textverarbeitungsdarstellung des Dokuments die zu erhaltende Inhaltsinformation ist, sondern nur eine Sequenz von Textabschnitten, die durch ASCII-Schriftzeichen dargestellt werden können. In diesem Fall könnte das OAIS entscheiden, die relevanten Textzeichen zu extrahieren und als Textdatei zu speichern. Das Inhaltsdatenobjekt würde in diesem Fall wahrscheinlich als Bitstream definiert werden, der aus diesen Zeichen gebildet wird. Die Repräsentationsinformationen wären dann die Beschreibung, wie dieser Bitstream als Zeichenfolge interpretiert werden kann, zusammen mit anderer zusätzlicher Information, die für ein angemessenes Verständnis des Textes notwendig ist.

4.2.1.4.2 Erhaltungsmetadaten

Zusätzlich zur Inhaltsinformation müssen die Archivinformationspakete Informationen beinhalten, welche das Vertrauen in, den Zugriff auf und den Kontext der Inhaltsinformation über eine unbestimmte Zeitspanne unterstützen. Der spezifische Satz an Informationsobjekten, die dafür benötigt werden, wird insgesamt als **Erhaltungsmetadaten** bezeichnet. Die Erhaltungsmetadaten müssen Informationen beinhalten, die für die angemessene Erhaltung der zugehörigen Inhaltsinformation notwendig sind. Besonders müssen sie den vergangenen und gegenwärtigen Zustand der Inhaltsinformation beschreiben, um deren eindeutige Identifizierung zu gewährleisten und sicherzustellen, dass sie nicht unwissentlich verändert wurde.

Diese Information ist typisch für alle Arten von Archiven und von traditionellen Archiven bereits klassifiziert worden. Allerdings müssen diese Klassendefinitionen für digitale Archive erweitert werden. Die folgenden Definitionen basieren weitgehend auf den Kategorien, die in dem Aufsatz „Preserving Digital Information“[16] besprochen wurden. Die Beziehung zwischen den Konzepten im OAIS-Referenzmodell und diesem Aufsatz werden in Anhang A[17] diskutiert. Tabelle 4-1 liefert Beispiele für verbreitete Typen von Inhaltsinformations-Typen.

Referenzinformation identifiziert – und beschreibt falls nötig – einen oder mehrere Mechanismen, mit denen der Inhaltsinformation Identifikatoren zugewiesen werden. Sie liefert auch Identifikatoren, die es externen Systemen erlauben, unzweideutig auf bestimmte Inhaltsinformation zu verweisen. Beispiele für solche Systeme umfassen Klassifizierungssysteme, Referenzsysteme und Registrierungssysteme. Im OAIS-Referenzmodell sind die meisten, wenn nicht sogar alle Information, auch in den Paketbeschreibungen enthalten, die den Endnutzern den Zugang zur gesuchten Inhaltsinformation ermöglichen.

Kontextinformation dokumentiert die Beziehungen der Inhaltsinformation zu ihrer Umgebung. Dazu gehört etwa, warum die Inhaltsinformation erzeugt wurde und wie sie sich zu anderswo vorhandenen Inhaltsinformationsobjekten verhält.

Provenienzinformation dokumentiert die Geschichte der Inhaltsinformation. Sie nennt den Ursprung oder die Quelle der Inhaltsinformation, alle Änderungen seit ihrer Entstehung und wer sie seitdem in Obhut hatte, und liefert damit eine Nachweiskette für die Inhaltsinformation. Das gibt künftigen Benutzern eine gewisse Zusage, wie zuverlässig die Inhaltsinformation ist, da es zum Nachweis der Authentizität beiträgt.

Die Provenienz kann als eine besondere Art von Kontextinformation angesehen werden.

Persistenzinformation stellt die Datenintegritätsprüfungen oder Validierungs-/ Verifizierungsschlüssel zur Verfügung, mit deren Hilfe sichergestellt wird, dass das bestimmte Inhaltsinformationsobjekt nicht undokumentiert verändert wurde. Persistenzinformation beinhaltet spezielle Kodierungsverfahren und Fehlererkennungsmethoden, die für Instanzen der Inhaltsobjekte spezifisch sind. Persistenzinformation beinhaltet nicht die Mechanismen zur Integritätserhaltung, die von den dem OAIS zugrundeliegenden Diensten bereitgestellt werden, nämlich der Fehlersicherung der Datenträger und der Gerätetreiber im ‚Archivspeicher'. Die Persistenzinformation kann aber Mindestanforderungen an die Servicequalität dieser Mechanismen definieren.

[16] Preserving Digital Information: Report of the Task Force on Archiving of Digital Information. Washington, D.C.: Commission on Preservation and Access, May 1996. <http://www.rlg.org/ArchTF/>

[17] Die Anhänge sind in der deutschen Übersetzung nicht enthalten.

Information über Zugriffsrechte identifiziert Zugriffsbeschränkungen für die Inhaltsinformation, einschließlich des rechtlichen Rahmens, Lizenzbedingungen und Zugriffskontrollen. Sie beinhaltet die Zugriffs- und Verbreitungsbedingungen, die in der Übergabevereinbarung festgelegt sind und die sowohl die Erhaltung (durch das OAIS) als auch die Benutzung (durch den Endnutzer) betreffen. Sie beinhaltet auch die Spezifikation für die Anwendung von Maßnahmen zur Rechteabsicherung.

Diese Klassifikationen stellen den minimalen Satz von Erhaltungsmetadaten dar, spezifizieren aber keine Datenstruktur.

Inhalts-informationstyp	Referenz	Provenienz	Kontext	Beständigkeit	Zugriffsrechte
Daten der Weltraumforschung	- Objekt Identifikator - Verweis auf Fachzeitschriften - Mission, Instrument, Titel, Satz von Merkmalen	- Instrumenten-Beschreibung - Projektleiter - Verarbeitungs-historie - Speicherungs- und Aufbewahrungs-historie - Sensor-beschreibung - Instrument - Instrumentenmodus - Entflechtungs-Abbildung - Spezifikation der Softwareschnitt-stellen - Informations-eigenschafts-beschreibung	- Kalibrierungs-geschichte - Verwandte Datensets - Mission - Finanzierungs-historie	- CRC - Prüfsumme - Reed-Solomon-Kodierung	- Identifikation der korrekt authorisierten vorgesehenen Zielgruppe (Zugriffskontrolle) - Erlaubnisnachweise für die Erhaltung und Verbreitung - Verweise zur Persistenz- und Provenienz-information (z.B. digitale Signaturen und Rechteinhaber)

Digitale Bibliotheks-kollektion	- Bibliographische Beschreibung
	- Persistenter Identifikator
		- Für gescannte Kollektionen:
		  - Metadaten über den Digitali-sierungs-prozess
		  - Verweis auf Master-version
		
		- Für originär digitale Publi-kationen:
		  - Verweis auf das digitale Original
		
		- Metadaten über den Erhaltungs-prozess:
		  - Verweis auf frühere Versionen des Kollektionsobjekts
		  - Änderungs-Geschichte
		  - Informations-eigenschafts-beschreibung
			- Verweis auf ähnliche Doku-mente in der Originalumge-bung zum Zeitpunkt der Veröffentlichung	- Digitale Signatur
				- Prüfsumme
				- Indikator der Authentizität
					- Rechtlicher Rahmen
					
					- Lizenzangebote
					
					- Spezifikation der Maßnahmen zur Rechtewahrung zur Auslieferungszeit
					
					- Erlaubnisnachweise für die Erhaltung und Verbreitung
					
					- Informationen über Wasserzeichen zum Zeitpunkt der Übergabe und Erhaltung
					
					- Verweise zur Persistenz- und Provenienz-information (z.B. digitale Signaturen und Rechteinhaber)

Softwarepaket	- Name - Autor/Urheber - Versionsnummer - Seriennummer	- Änderungs-geschichte - Registrierung - Urheberrecht - Informations-eigenschafts-beschreibung	- Hilfedatei - Benutzerhandbuch - Verwandte Soft-ware - Sprache	- Zertifikat - Prüfsumme - Verschlüsselung - CRC	- Vorgesehene Zielgruppe - Rechtlicher Rahmen - Lizenzangebote - Spezifikation der Maßnahmen zur Rechtewahrung zur Auslieferungszeit - Verweise zur Persistenz - und Provenienz-information (z.B. digitale Signaturen und Rechteinhaber)

Tabelle 4-1: Beispiele für Erhaltungsmetadaten

Das OAIS muss explizit entscheiden, wie die Inhaltsinformation genau definiert ist, damit es sicherstellen kann, dass es auch die benötigten Erhaltungsmetadaten hat, um die Inhaltsinformation zu erhalten. Sobald die Inhaltsinformation festgelegt ist, ist es möglich, die Erhaltungsmetadaten zu bestimmen.

4.2.1.4.3 Verpackungsinformation

Die **Verpackungsinformation** ist die Information, die die Komponenten des Pakets entweder tatsächlich oder logisch in einer identifizierbaren Einheit auf einem bestimmten Datenträger zusammenhält oder verknüpft. Zum Beispiel: Falls die Inhaltsinformation und die Erhaltungsmetadaten als der Inhalt bestimmter Dateien in einer TAR-Datei identifiziert werden, dann können die Verpackungsinformationen den Namen der TAR-Datei enthalten und die Tatsache, dass es sich um eine TAR-Datei handelt, inklusive Details über spezielle Kodierungen. Wenn aber die Inhaltsinformationen und Erhaltungsmetadaten Dateien auf einer CD-Rom sind, dann kann die Verpackungsinformation die Datenträger- und Dateistruktur auf der CD-ROM gemäß ISO 9660 enthalten. Diese Auswahl wird von jedem Archiv in seinen Definitionen oder Richtlinien individuell getroffen. Die Verpackungsinformation muss von einem OAIS nicht unbedingt erhalten werden, weil sie nichts zu den Inhaltsinformation oder den Erhaltungsmetadaten beiträgt. Trotzdem kann es Fälle geben, in denen ein OAIS verpflichtet sein kann, die ursprüngliche Übergabe exakt wiederzugeben. In diesem Fall wird die Inhaltsinformation so definiert, dass sie alle übergebenen Bits enthält.

Das OAIS sollte vermeiden, Erhaltungsmetadaten oder Inhaltsinformation einzig in den Namenskonventionen für Verzeichnisse oder Dateien bereit zu halten. Diese Strukturen werden höchst wahrscheinlich als Verpackungsinformation verwendet. Verpackungsinformationen bleiben aber nicht bei allen digitalen Migrationen erhalten. Information, die nur in Dateinamen oder in den Namen der Inhaltsverzeichnisse enthalten sind, können verloren gehen, wenn die Verpackungsinformation geändert wird. Das Thema Verpackungsinformation spielt eine wichtige Rolle bei der Migration von Information innerhalb eines OAIS auf neuere Datenträger. Dieses Thema wird detailliert in Abschnitt 5 dieses Dokuments behandelt.

4.2.1.4.4. Erschließungsinformationen

Die Informationsobjekte, die bisher in diesem Abschnitt beschrieben worden sind, liefern die notwendige Information, um die Langzeiterhaltung zu ermöglichen. Zusätzlich zum Erhalt der Information muss das OAIS den Endnutzern die Möglichkeit bieten, die gewünschte Information zu finden, analysieren und bestellen. Das wird durch einen Spezialfall des Informationsobjekts erreicht, der Erschließungsinformation genannt wird und Daten enthält, die als Eingabe für Dokumente oder Anwendungen dienen, die Zugriffshilfen genannt werden. Die Erschließungsinformationen werden üblicherweise von den Inhaltsinformationen und den Erhaltungsmetadaten abgeleitet. Man kann sie als einen Index betrachten, der über die damit verbundenen Zugriffshilfen einen effizienten Zugriff auf die entsprechenden Informationspakete ermöglicht. Zugriffshilfen sind Dokumente oder Programme, die benutzt werden können, um Information im OAIS ausfindig zu machen, auszuwerten, abzurufen oder zu bestellen.

4.2.2. Logisches Modell der Information in einem Offenen Archiv-Informations-System (OAIS)

Der vorherige Unterabschnitt definierte die Typen von Informationsobjekten, die ein OAIS für die Langzeiterhaltung von Information und den Zugriff der vorgesehenen Zielgruppe auf die archivierte Information braucht. Dieser Unterabschnitt verwendet die Beschreibungen dieser Informationsobjekte, um die konzeptionellen Informationsstrukturen zu modellieren, die zur Erfüllung dieser Funktionen gebraucht werden. Die in diesem Unterabschnitt präsentierten Modelle sollen keine Implementierung vorwegnehmen, sondern vielmehr die Beziehungen zwischen den verschiedenen Arten von Information näher beleuchten, die im Archivierungsprozess gebraucht werden.

4.2.2.1 Informationspaket

Die konzeptionelle Struktur, die die Langzeiterhaltung von Information unterstützt, ist das Informationspaket. Ein Informationspaket ist ein Container, der zwei Arten von Informationsobjekten enthält: die Inhaltsinformationen und die Erhaltungsmetadaten. Das Informationspaket kann verbunden werden mit zwei anderen Arten von Informationsobjekten: mit der Verpackungsinformation und der Paketbeschreibung. Es gibt verschiedene Arten von Informationspaketen, die innerhalb des Archivierungsprozesses verwendet werden. Diese Informationspakete können benutzt werden, um den Bestand des OAIS zu strukturieren und zu speichern; um die benötigte Information vom Produzenten zum OAIS zu transportieren; oder um die angeforderten Information zwischen dem OAIS und den Endnutzern zu transportieren. Jede dieser Funktionen erfordert unterschiedliche Informationen. Das UML-Diagramm in Abbildung 4-13 illustriert das Konzept eines Informationspakets. Es zeigt, dass ein Informationspaket kein oder ein Inhaltsinformationsobjekt enthält, kein oder mehrere Erhaltungsmetadaten-Objekte, und mit genau einer Verpackungsinformation verbunden ist, die das Informationspaket identifiziert und begrenzt. Das Informationspaket ist außerdem verbunden mit einer oder mehreren Paketbeschreibungen, die das Inhaltsobjekt beschreiben, um den effizienten Zugriff auf diese zu ermöglichen.

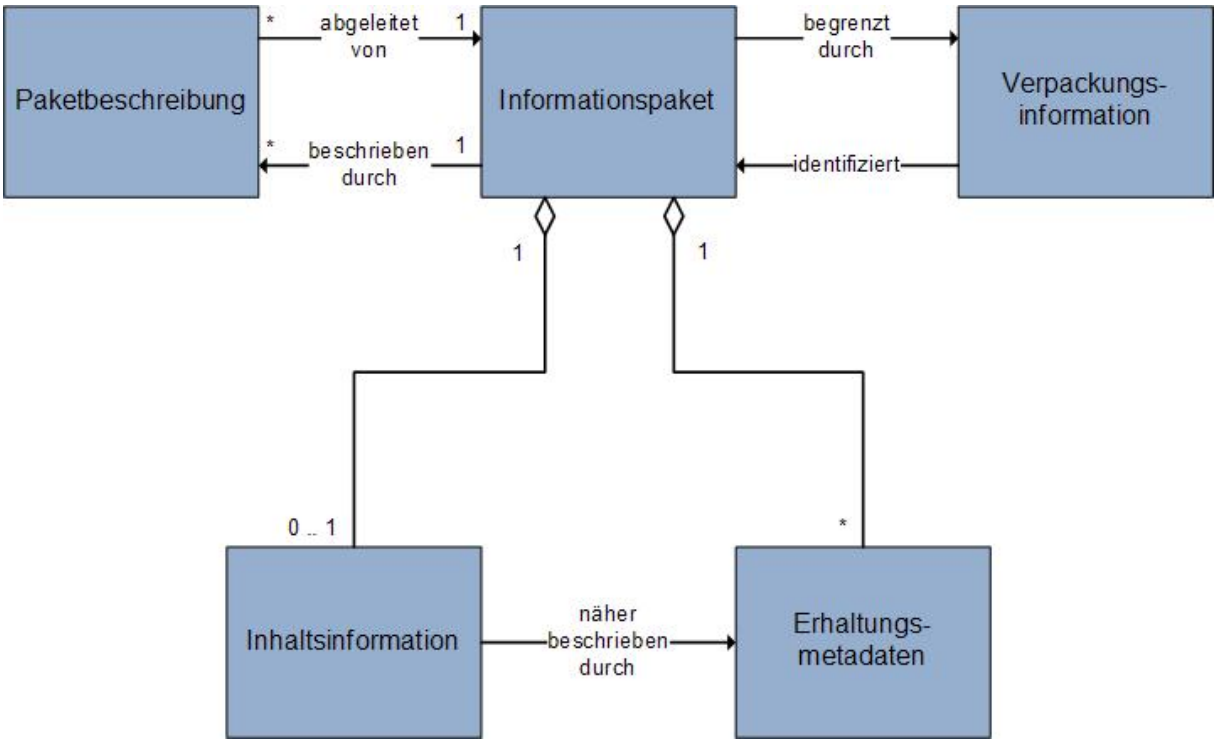

Abbildung 4-13: Inhalte eines Informationspakets

4.2.2.2 Typen von Informationspaketen

Es gibt drei Untertypen von Informationspaketen: Übergabeinformationspakete (SIPs), Archivinformationspakete (AIPs) und Auslieferungsinformationspakete (DIPs). Die Definitionen dieser Pakettypen in Abschnitt 2 basieren auf der Funktion im Archivierungsprozess, die das Paket benutzt, und die Überführung von einem Paket in das nächste im Lauf des Archivierungsprozesses. Die Taxonomie der Pakettypen wird in Abbildung 4-14 gezeigt.

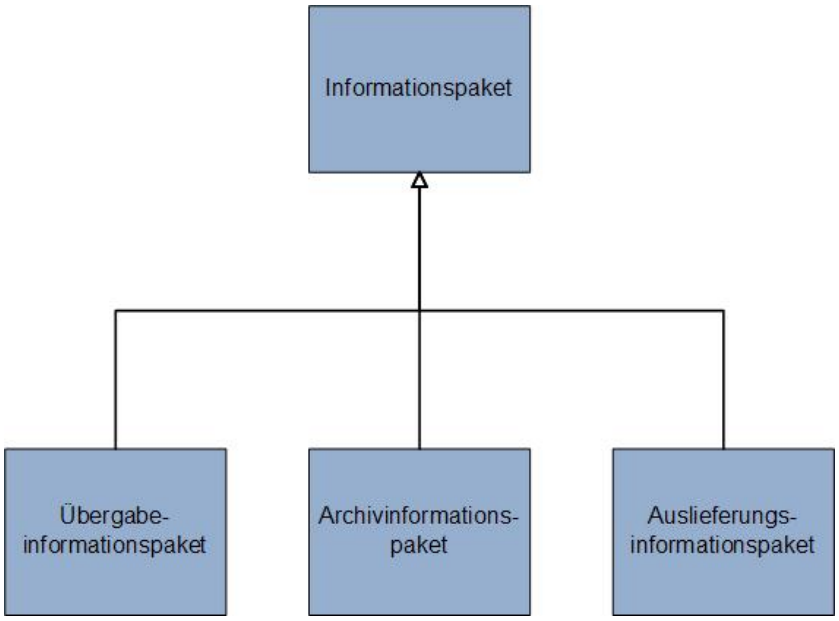

Abbildung 4-14: Taxonomie der Informationspakete

Es muss unterschieden werden zwischen einem Informationspaket, das vom OAIS erhalten wird, und den Informationspaketen, die an das OAIS übergeben und vom OAIS ausgeliefert werden. Diese verschiedenen Varianten sind ein Abbild der Realität, dass gewisse Übergaben an ein OAIS nur unzureichende Repräsentationsinformationen oder Erhaltungsmetadaten enthalten, um den Anforderungen des OAIS an die Erhaltung zu genügen. Außerdem kann sich die interne Struktur der übernommenen Pakete stark von der im OAIS bevorzugten Struktur unterscheiden. Schließlich wird das OAIS möglicherweise Informationen an die Endnutzer herausgeben, die nicht alle Repräsentationsinformationen oder Erhaltungsmetadaten zu den ausgelieferten Inhaltsinformationen enthalten. Diese Varianten werden als Übergabeinformationspakete (SIPs), Archivinformationspakete (AIPs) und Auslieferungsinformationspakete (DIPs) bezeichnet. Auch wenn es sich bei allen drei Varianten um Informationspakete handelt, unterscheiden sie sich in ihrem obligatorischen Inhalt und in der Vielzahl der Beziehungen zwischen den enthaltenen Klassen.

Das **Übergabeinformationspaket (SIP)** ist das Paket, das dem OAIS vom Produzenten geschickt wird. Seine Form und sein genauer Inhalt werden üblicherweise zwischen Archiv und Produzent ausgehandelt. Die meisten SIPs werden einige Inhaltsinformationen sowie einige Erhaltungsmetadaten enthalten, aber unter Umständen sind mehrere SIPs notwendig, um einen kompletten Satz von Inhaltsinformationen und zugehörigen Erhaltungsmetadaten zu erhalten. Die Inhaltsinformationen und die Erhaltungsmetadaten haben beide jeweils zugehörige Repräsentationsinformationen, und falls mehrere SIPs involviert sind, die dieselben Repräsentationsinformationen verwenden, werden diese Repräsentationsinformationen wahrscheinlich nur einmal an das OAIS übermittelt. Weil manche Arten von Erhaltungsmetadaten auf viele SIPs aus derselben Quelle zutreffen, kann es außerdem sein, dass solche Erhaltungsmetadaten in einem separaten SIP ohne Inhaltsinformationen zur Verfügung gestellt werden. Die Verpackungsinformation ist immer in irgendeiner Form vorhanden.

Die Erschließungsinformationen, die mit einem SIP verbunden sind, werden wahrscheinlich vor der Übermittlung des SIPs an das OAIS geliefert, können aber auch zu jedem beliebigen Zeitpunkt geliefert werden. Es kann sich dabei um nicht mehr als eine Beschreibung in Textform mit einem Namen oder Titel handeln, transportiert von der Verpackungsinformation, anhand derer das SIP erkannt wird.

Im OAIS werden ein oder mehrere SIPs zur Erhaltung in ein oder mehrere Archivinformationspakete umgewandelt. Das AIP enthält einen kompletten Satz von Erhaltungsmetadaten zur dazugehörigen Inhaltsinformation. Das AIP kann auch eine Kollektion von anderen AIPs enthalten; diese Variante wird weiter unten in diesem Unterabschnitt diskutiert und modelliert. Die Verpackungsinformation des AIPs muss den internen Standards des OAIS entsprechen und kann sich während der Verwaltung durch das OAIS ändern. Die zu einem AIP gehörenden Erschließungsinformationen können sehr ausführlich sein und werden vom Archiv so verwaltet, dass die Endnutzer die gewünschten Inhaltsinformationen finden und bestellen können.

Als Antwort auf eine Bestellung übermittelt das Archiv das ganze oder einen Teil eines AIPs an den Endnutzer in Form eines Auslieferungsinformationspakets (DIP). Das DIP kann auch Kollektionen von AIPs sowie die kompletten oder Teile der Erhaltungsmetadaten enthalten. Die Verpackungsinformation ist immer in irgendeiner Form vorhanden, so dass der Endnutzer die bestellte Information deutlich unterscheiden kann. Die Verpackungsinformation wird je nach Auslieferungsmedium und Endnutzerbedürfnissen variieren. Die Erschließungsinformationen, die mit einem DIP verbunden sind, können gleichzeitig mit, aber auch vor oder nach dem Transfer des DIPs übermittelt werden. Der Zweck der Erschließungsinformationen ist es, dem Endnutzer genügend Information an die Hand zu geben, damit er das DIP von anderen, möglicherweise ähnlichen Paketen unterscheiden kann. Es kann sich dabei um nicht mehr als eine Beschreibung in Textform mit einem Namen oder Titel handeln, die von der Verpackungsinformation transportiert werden, anhand derer das DIP erkannt werden kann.

Obwohl die Umsetzung eines AIPs von Archiv zu Archiv variieren kann, bleibt die allgemeingültige Definition eines AIPs als Container bestehen, der alle Information enthält, die für die Langzeiterhaltung sowie für den Zugriff auf Archivbestände gebraucht werden. Das in 4.2.2.3 dargestellte Modell eines AIPs soll als Referenz dienen, um die Arten von Information festzulegen, die für Langzeiterhaltung und -zugriff gebraucht werden.

Die genauen Inhalte der SIPs und DIPs und ihr Verhältnis zu den entsprechenden AIPs hängen ab von den Vereinbarungen zwischen dem Archiv und seinen Produzenten und Endnutzern. Das Modell für diese beiden Pakete ist das gleiche wie für das Informationspaket, das in Abbildung 4-15 gezeigt wird, sowohl was den obligatorischen Inhalt als auch die Vielzahl der Beziehungen zwischen den einzelnen Bestandteilen angeht. Das Thema Transformation von SIP in AIP und von AIP in DIP wird in 4.3 näher besprochen.

4.2.2.3 Das Archivinformationspaket

Das Archivinformationspaket (AIP), das in Abbildung 4-15 modelliert wird, ist ein Spezialfall eines Informationspakets. Das AIP ist definiert als eine präzise Art, auf einen Satz von Informationen zu verweisen, der im Prinzip alle notwendigen Eigenschaften für eine permanente, oder unbegrenzte, Langzeiterhaltung eines bestimmten Informationsobjekts hat. Das AIP ist selbst ein Informationsobjekt, das ein Container für andere Informationsobjekte ist. Innerhalb des AIPs befindet sich das bestimmte Informationsobjekt, das hier "Inhaltsinformation" genannt wird.

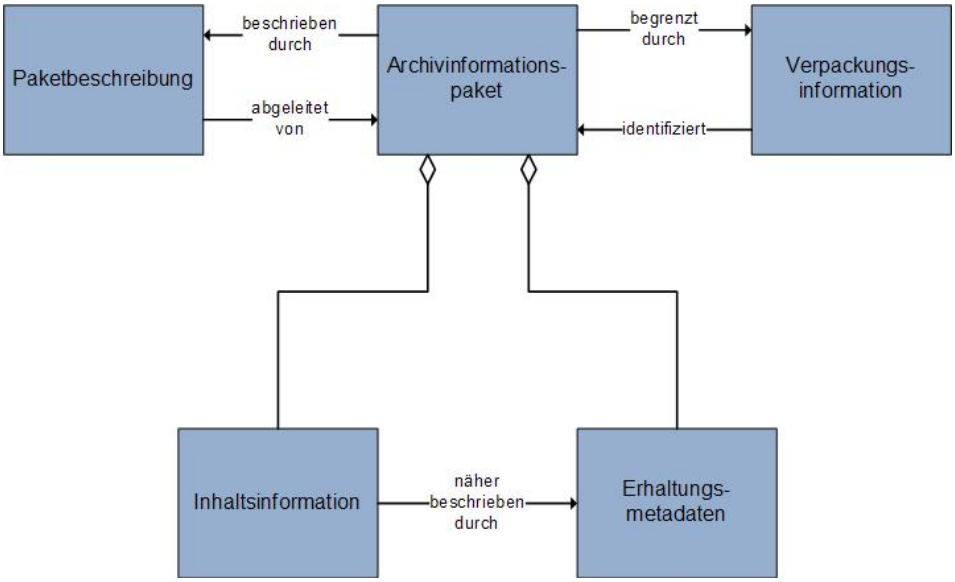

Abbildung 4-15: Archivinformationspaket (AIP)

Innerhalb des AIPs befindet sich außerdem ein Informationsobjekt, das "Erhaltungsmetadaten" genannt wird. Diese Erhaltungsmetadaten enthalten zusätzliche Information über die Inhaltsinformation und werden gebraucht, um die Inhaltsinformation für eine unbestimmt langfristige Zeit aussagekräftig zu machen.

Die Ansprüche an die Erhaltungsmetadaten in einem AIP sind viel strenger als die Ansprüche an die Erhaltungsmetadaten in einem allgemeinen Informationspaket. Während Erhaltungsmetadatenobjekte in einem Informationspaket nicht obligatorisch vorkommen, müssen alle Arten von Erhaltungsmetadaten in einem AIP vorhanden sein (s. Abbildung 4-16). Wie die einzelnen Erhaltungsmetadaten befüllt werden, ist Sache des jeweiligen Archivs.

Zum Beispiel kann in gewissen OAIS-Beständen die Angabe, dass der Urheber der Inhaltsinformation unbekannt ist, als Provenienzinformation ausreichend sein, während es in anderen OAIS-Beständen unabdingbar ist, die Provenienz ausführlicher zu recherchieren.

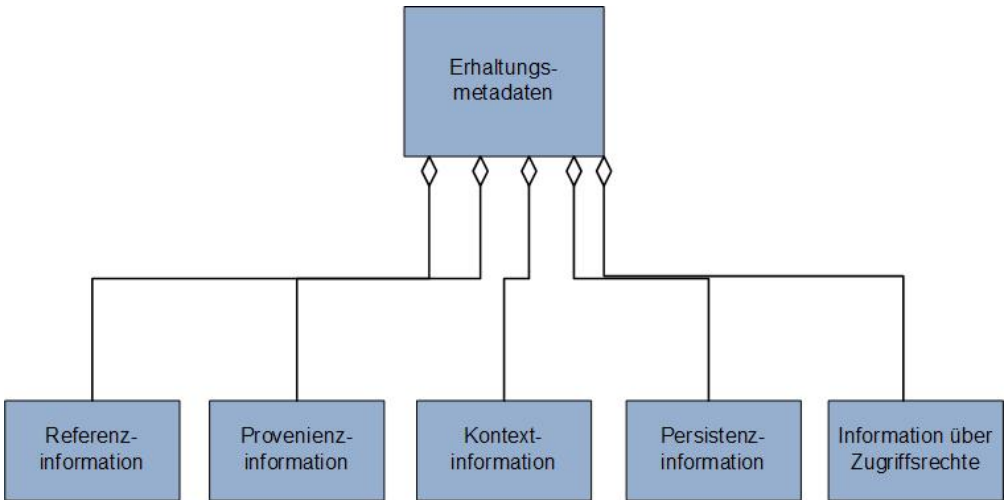

Abbildung 4-16: Erhaltungsmetadaten

Das AIP wird abgegrenzt und identifiziert durch die Verpackungsinformation. Die Verpackungsinformation kann tatsächlich als Struktur auf dem Datenträger vorhanden sein, der das AIP enthält, oder sie kann virtuell in der Funktion ‚Archivspeicher' des OAIS enthalten sein. In jedem Fall müssen die Funktionen zur Abgrenzung und zur internen Identifikation in einem OAIS präzise definiert sein.

Zu jedem AIP gehört eine strukturierte Form von Erschließungsinformationen, die **Paketbeschreibung** genannt wird. Sie ermöglicht es dem Endnutzer, potenziell interessante Informationen zu finden, diese zu analysieren und die gewünschte Information zu bestellen. Die für eine Zugriffshilfe gebrauchte Information wird **Begleitbeschreibung** genannt. Eine einzige Paketbeschreibung kann verschiedene Begleitbeschreibungen enthalten, abhängig von der Zahl der verschiedenen Zugriffshilfen, welche die zusammengehörigen Inhaltsinformation und Erhaltungsmetadaten suchen, anzeigen, abfragen oder bestellen können. Abbildung 4-17 zeigt ein UML-Diagramm der Paketbeschreibung und der Zugriffshilfe.

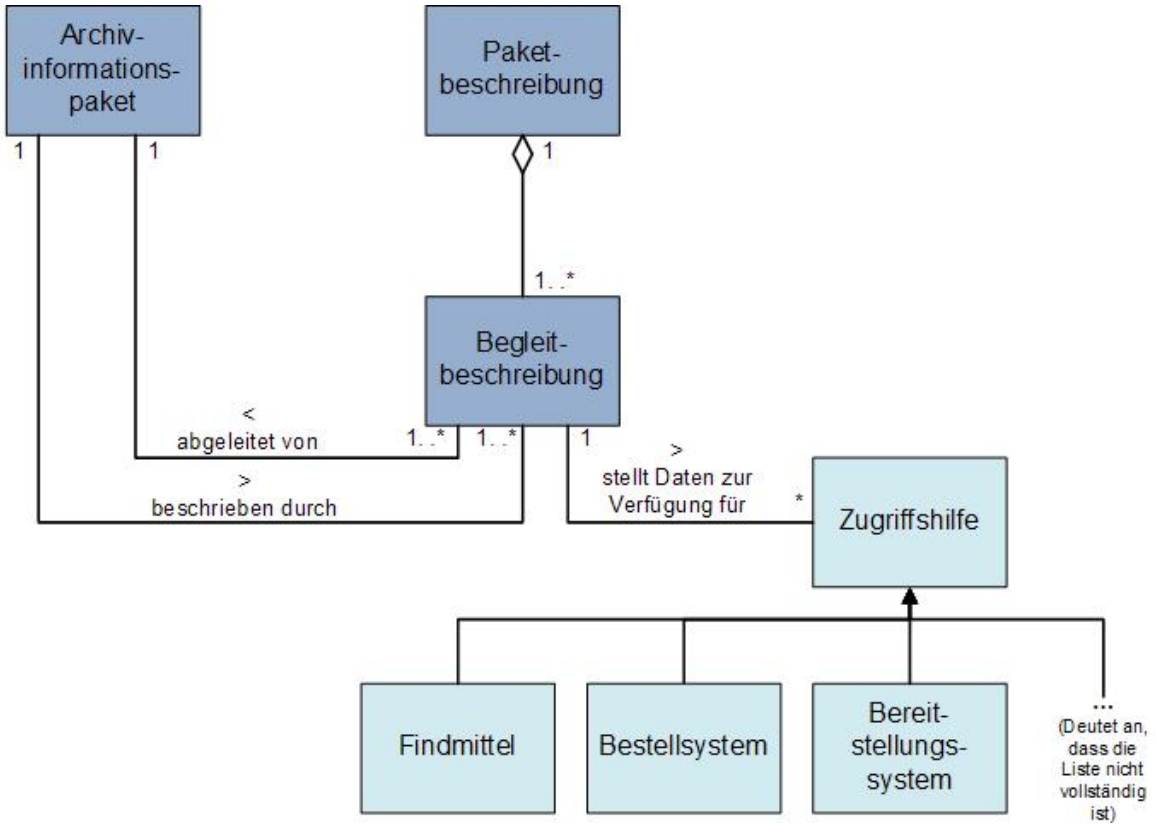

Abbildung 4-17: Paketbeschreibung

Die Paketbeschreibung muss eine Begleitbeschreibung enthalten, die Daten für ein Bereitstellungssystem liefert, mit dem autorisierte Benutzer die von der Paketbeschreibung beschriebenen Inhaltsinformationen und Erhaltungsmetadaten finden können. Dieses Bereitstellungssystem ist normalerweise Teil des Funktionsbereichs ‚Archivspeicher'. Es übersetzt den eindeutigen Identifier, der dem AIP zur Identifikation vom OAIS zugewiesen worden ist, in ein Set von Prozessen und Dateinamen, die zum Auffinden des AIPs im Dateiverwaltungssystem des ‚Archivspeichers' gebraucht werden, und gibt dann die Inhaltsinformation und Erhaltungsmetadaten des angefragten AIPs aus. Gegenwärtig dürfen in den meisten Archiven nur archivinterne Prozesse und das Betriebspersonal diese Zugriffshilfe verwenden. Da jedoch der technologische Fortschritt die Rechenleistung des Archivs und die Bandbreite zwischen Archiv und Benutzern vergrößert, können Zugriffsmethoden wie „inhaltsbezogene Anfragen" oder „Data Mining" dem Benutzer direkten Lesezugriff auf die Inhaltsinformation erlauben.

Die Paketbeschreibung kann außerdem eine beliebige Anzahl Begleitbeschreibungen enthalten, von denen jede einzelne Daten für eine oder mehrere Zugriffshilfen enthält. Zwei zusätzliche Untertypen von Zugriffshilfen sind das **Findmittel** und das **Bestellsystem**.

Ein **Findmittel** ist eine Anwendung, die dem Endnutzer dabei hilft, Informationen zu finden. Ein einzelnes AIP kann mehrere Begleitbeschreibungen haben, die die Inhaltsinformation mittels unterschiedlicher Technologien beschreiben.

Ein **Bestellsystem** ist eine Anwendung, die den Endnutzer dabei unterstützt, die gewünschten AIPs zu bestellen und die Kosten dafür zu ermitteln. Ein Bestellsystem ermöglicht es dem Benutzer auch, Transformationen zu spezifizieren, welche vor der Auslieferung auf die AIPs angewendet werden sollen. Dazu können Transformationen der Datenobjekte in eine Untermenge, in eine kleinere Zusammenstellung oder in ein anderes Format gehören. Eine Transformation kann auch bedeuten, dass die Erhaltungsmetadaten des AIPs vor der Auslieferung verändert werden.

Die Paketbeschreibung ist für die Langzeiterhaltung der Inhaltsinformation nicht erforderlich, aber sie wird gebraucht, um die Bestände eines Archivs sichtbar und zugreifbar zu machen. Die Inhalte der Paketbeschreibung hängen in hohem Maße von der Struktur der Inhaltsinformation und der Erhaltungsmetadaten ab, die sie beschreibt. Die Verwendung und die Typen von Paketbeschreibungen sind im 4.2.2.4 ausführlicher definiert.

Abbildung 4-18 zeigt detailliert ein Archivinformationspaket mit Erhaltungsmetadaten und Inhaltsinformationen. Alle „Enthält“-Beziehungen, die in diesem Unterabschnitt diskutiert werden, sind logische Beinhaltungs-Beziehungen. Diese Art von „Enthält“-Beziehung kann physisch sein oder mittels Verweis auf ein anderes Objekt im Speicher erreicht werden, so dass ein AIP nicht zwingend eine einzelne Datei ist.

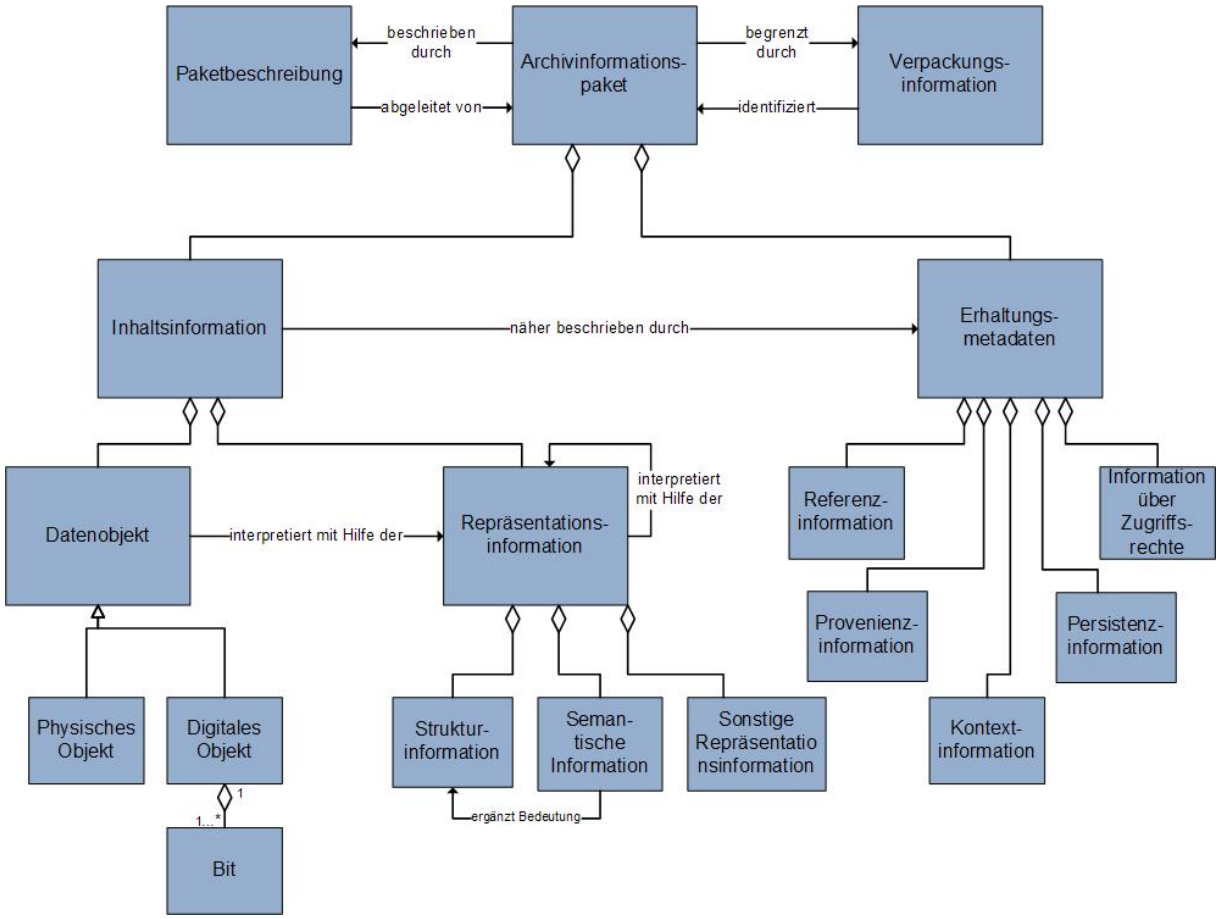

Abbildung 4-18: Detaillierte Ansicht eines Archivinformationspakets

4.2.2.4 Spezialfälle des AIPs und der Paketbeschreibung

In diesem Unterabschnitt werden zwei Spezialfälle des AIPs diskutiert, die **Archivinformationseinheit (AIU)** und die **Archivinformationskollektion (AIC)**. Abbildung 4-19 zeigt ein UML-Diagramm dieser beiden Spezialfälle. AIU und AIC sind beide Untertypen des AIPs und enthalten somit Konstrukte, die Langzeiterhaltung und Endnutzerzugriff ermöglichen. Die AIU ist der Typus, der für die Erhaltung von Inhaltsinformation verwendet wird, die nicht in andere AIPs aufgeteilt werden kann. Die AIC ordnet einen Satz von AIPs (AIUs und andere AICs) nach einer thematischen Hierarchie, die den flexiblen und effizienten Zugriff durch die Endnutzer-Gemeinschaft unterstützen kann. Vom Konzept her sind alle AIPs einer AIC in der Inhaltsinformation dieser AIC enthalten. Den Unterschied zwischen AIUs und AICs macht die Komplexität ihrer Inhaltsinformation und der damit verbundenen Paketbeschreibungen und Verpackungsinformationen aus. Dieses Referenzmodell sieht die Unterschiede in der

Inhaltsinformation und der damit verbundenen Paket- und Beschreibungsfunktionalitäten von AIU und AIC als hinreichend komplex und aufeinander bezogen an, um die Einordnung in verschiedene Klassen zu rechtfertigen.

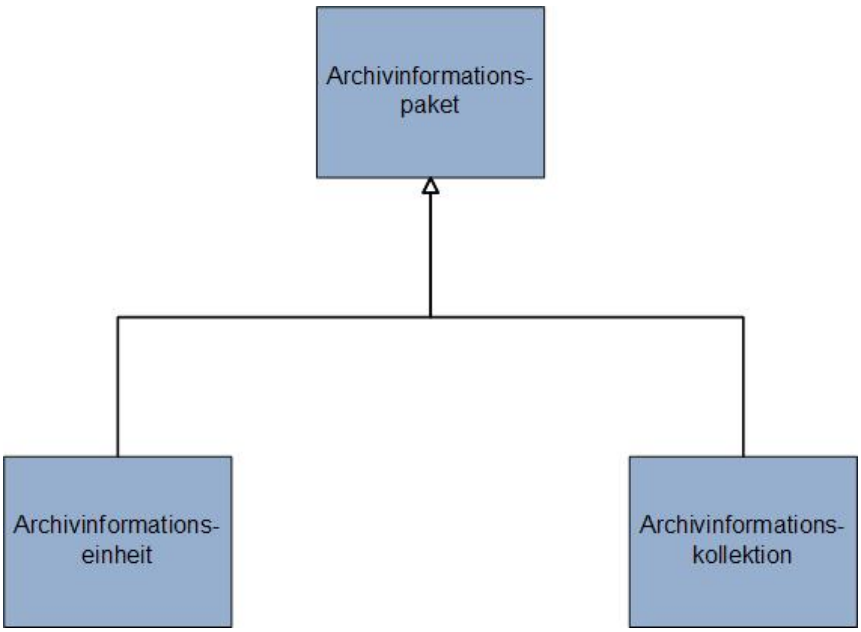

Abbildung 4-19: Spezialfälle des AIPs

Vom Standpunkt des ‚Zugriffs' aus beginnen die Unterschiede zwischen AICs und AIUs durch neue Möglichkeiten der Unterteilung und Bearbeitung zu verschwimmen. Bisher als atomar betrachtete Inhaltsobjekte können jetzt je nach dem gewählten Unterteilungsgrad als Container interpretiert werden, die eine große Vielfalt von Inhalten enthalten. Im Extremfall ist es sogar möglich, dass die Inhaltsinformation einer AIU überhaupt nicht als physikalische Einheit existiert. Die Inhaltsinformation könnte aus verschiedenen Eingabe-Dateien (oder aus Verweisen auf die AIPs, die diese Dateien enthalten) und aus einem Algorithmus bestehen, der diese Dateien benutzt, um das gewünschte Datenobjekt zu erzeugen.

Vom Standpunkt der Informationserhaltung aus bleibt die Unterscheidung zwischen AIU und AIC klar. Eine AIU hat nur ein einzelnes Inhaltsinformationsobjekt, das mit genau einem Satz von Erhaltungsmetadaten beschrieben wird. Die Inhaltsinformation einer AIC hingegen besteht aus einer Kollektion von anderen AICs und AIUs, die jeweils eigene Erhaltungs-metadaten haben. Zusätzlich hat diese AIC eigene Erhaltungsmetadaten, die die Kriterien für die Kollektion und deren Prozess beschreiben.

Es gibt zwei Spezialfälle der Paketbeschreibung, die Archiveinheits-Beschreibung und die Kollektionsbeschreibung. Das UML-Klassendiagramm in Abbildung 4-20 illustriert diese Spezialfälle). Der Unterschied zwischen diesen beiden Klassen beruht auf den Funktionen, die für den effizienten Zugriff auf den Inhalt einer AIU beziehungsweise auf die in einer AIC enthaltenen AIPs gebraucht werden.

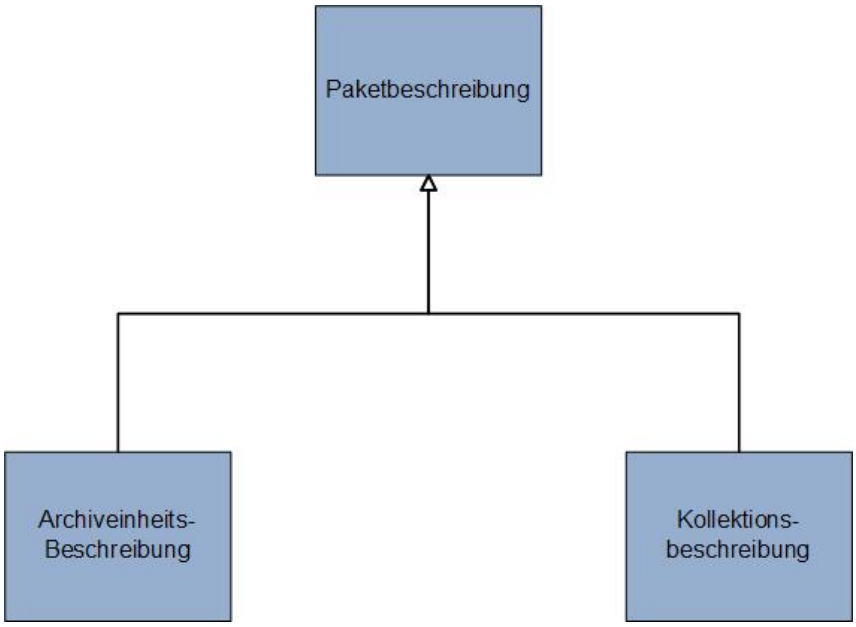

Abbildung 4-20: Spezialfälle des Pakets

Zum besseren Verständnis dieser Konstrukte wird in den nächsten beiden Unterabschnitten dieses Dokuments das Beispiel einer Firma beschrieben, die ein OAIS für digitale Versionen von Filmen einrichtet. Dieses Beispiel konzentriert sich auf die Information, die in den Konstrukten eines AIPs enthalten ist. Unterabschnitt 4.3 illustriert detaillierter die Transformation von Information und die Datenflüsse in einem OAIS.

4.2.2.5. Archivinformationseinheit (AIU)

Archivinformationseinheiten können als die „Atome“ der Information betrachtet werden, die ein Archiv aufbewahren soll. Eine einzelne AIU enthält genau ein Inhaltsinformations-Objekt (das aus mehreren Dateien bestehen kann) und genau einen Satz von Erhaltungsmetadaten. Es steht dem Archiv frei, wie eine AIU konstruiert sein soll. Insbesondere braucht eine AIU nicht eine einzelne Datei zu sein. Wenn ein Informationsobjekt ins OAIS übernommen wird, wird eine **Archiveinheits-Beschreibung** erzeugt, ein Untertypus der Paketbeschreibung. Dazu wird Information aus der Inhaltsinformation und den Erhaltungsmetadaten extrahiert und OAIS-spezifische Information wie etwa ein eindeutiger Identifikator hinzugefügt. Die AIU ist in Abbildung 4-21 illustriert.

Im Beispiel eines OAIS für digitale Filme kann die AIU für einen einzelnen Film als drei Objekte aufgefasst werden: eines enthält den Binärcode des Films in einem proprietären Format, ein anderes enthält die Repräsentationsinformationen für die Interpretation des proprietären Formats (diese zwei Objekte bilden die Inhaltsinformation) und das dritte enthält Fakten über den Film wie das Entstehungsdatum, Darsteller, Regisseur, Produzenten, Fortsetzungen, Studio sowie eine Prüfsumme zur Sicherstellung der Integrität des digitalen Films (Erhaltungsmetadaten). Weil das OAIS-Referenzmodell von keiner bestimmten Implementierung abhängig ist, könnte jedes dieser Objekte als eine Datei oder als mehrere Dateien implementiert werden. Diese implementationsabhängigen Arten von Information sind in der Verpackungsinformation enthalten. Wird ein Film in ein OAIS übernommen, kann eine Archiveinheits-Beschreibung für ein Bestellsystem erstellt werden, indem man Information aus der Inhaltsinformation und den Erhaltungsmetadaten extrahiert und an die eindeutige Bestellinformation anhängt.

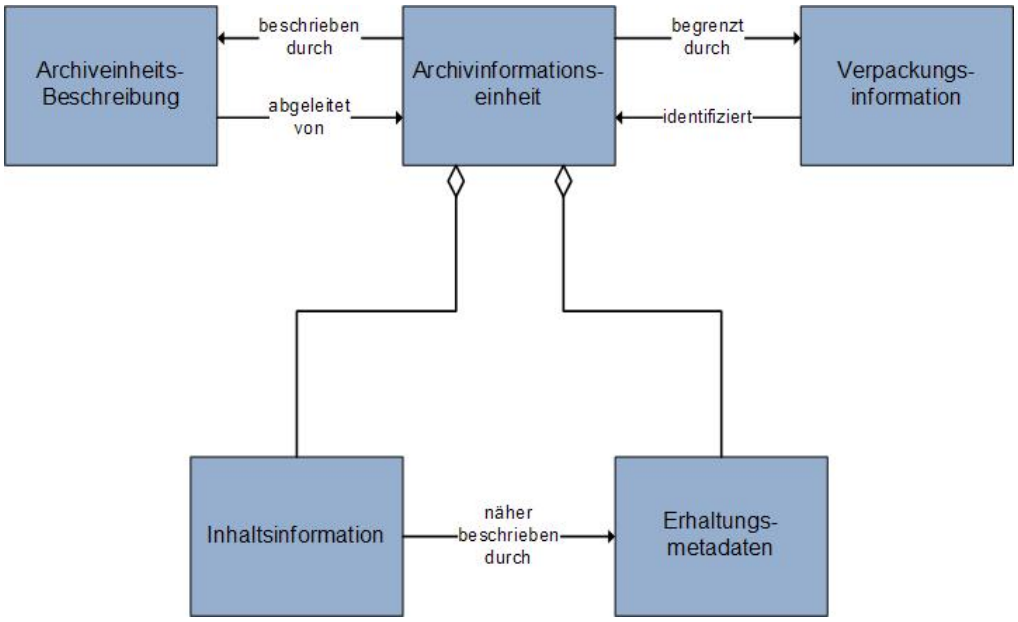

Abbildung 4-21: Archiveinheit (AIU)

4.2.2.6. Archiveinheits-Beschreibung

Die Archiveinheits-Beschreibung ist ein Spezialfall der Paketbeschreibung, die immer einen Satz von Begleitbeschreibungen enthält, von denen jede die Inhaltsinformation der AIU vom Standpunkt einer bestimmten Zugriffshilfe aus beschreibt. Abbildung 4-22 ist ein UML-Diagramm, das die Inhalte einer Archiveinheits-Beschreibung veranschaulicht.

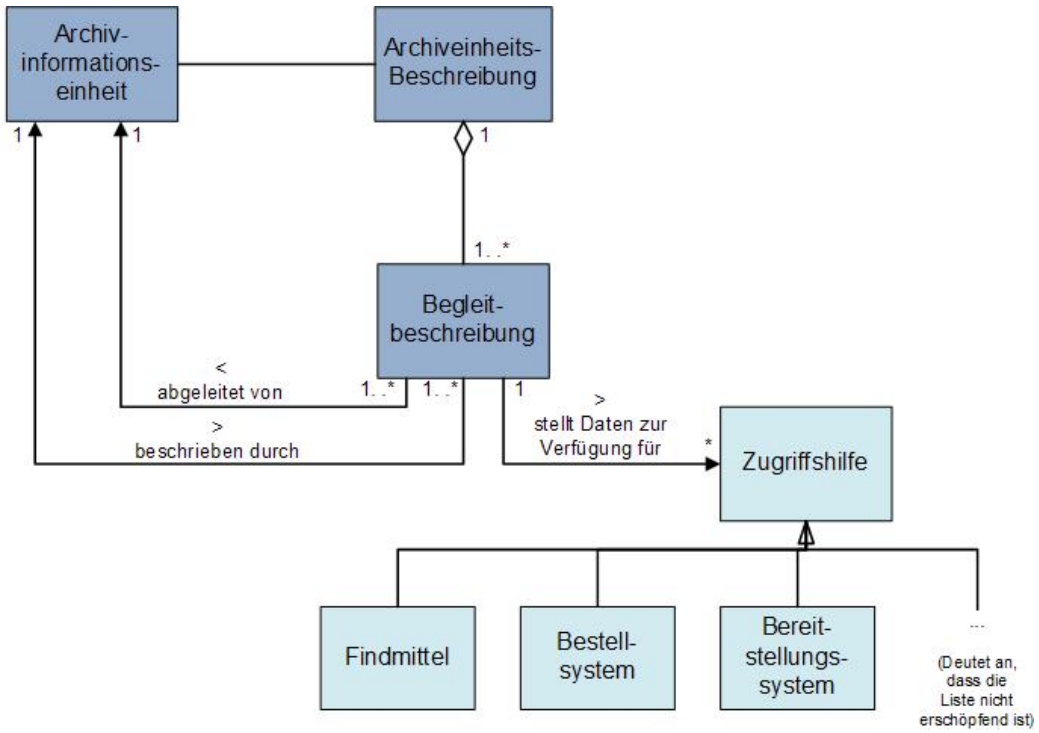

Abbildung 4-22: Archiveinheits-Beschreibung

Alle Archiveinheits-Beschreibungen müssen eine Begleitbeschreibung für ein Bereitstellungssystem enthalten, das es autorisierten Benutzern ermöglicht, die AIU im ‚Archivspeicher' zu finden, die durch die Archiveinheits-Beschreibung beschrieben wird. Diese

Beschreibung umfasst den eindeutigen Identifikator, der dem AIP durch den ‚Archivspeicher' während des Übernahmeprozesses zugeteilt worden ist.

Ein wichtiger Typ von Zugriffshilfen ist das Findmittel. Ein Findmittel ist eine Anwendung, die den Endnutzer bei der Recherche unterstützt. Eine einzelne AIU kann über eine Vielzahl von Begleitbeschreibungen verfügen, die unter Verwendung verschiedener Techniken die Inhaltsinformation beschreiben. Wenn neue Techniken zur Extraktion und Präsentation von Beschreibungen verfügbar werden, kann sich ein Archiv entscheiden, die Archiveinheits-Beschreibungen seiner AIUs zu aktualisieren und eine neue Begleitbeschreibung hinzuzufügen, welche die neue Technik nutzt, um die AIU besser zu beschreiben.

In unserem Beispiel-OAIS für digitale Filme gibt es zu Beginn vielleicht eine Begleitbeschreibung in Form einer Freitext-Beschreibung eines Films, eine andere in Form eines Fünf-Minuten-Clips und eine dritte als Zeile in einer relationalen Datenbank, die Sammler zum Auffinden interessanter Filme verwenden. Nach einer bestimmten Betriebszeit des Archivs wird dann vielleicht eine Technik entwickelt, die komprimierte digitale Filme bereitstellt, indem jedes 10. Bild aufgezeichnet wird. Der Archivar kann dann beschließen, eine zusätzliche Art der Begleitbeschreibung zu erstellen, für die die Ergebnisse dieser neuen Technik verwendet werden. Falls gewünscht kann der Benutzer jede AIU des Archivs auf diese Art komprimieren lassen und so eine neue Begleitbeschreibung für jeden Film im Archiv erstellen – oder aber diese Begleitbeschreibung wird einfach für alle neuen AIUs angehängt, die ins OAIS übernommen werden.

Eine weitere wichtige Klasse von Begleitbeschreibungen liefert die Daten für Bestellsysteme, die es den Endnutzern ermöglichen, die Kosten einer Bestellung zu ermitteln und AIUs zu bestellen. Die Bestellsysteme ermöglichen es den Benutzern außerdem, Transformationen zu spezifizieren, die auf eine AIU vor der Auslieferung angewendet werden müssen. Diese Transformationen können die Umwandlung von Datenobjekten beinhalten, etwa in eine Untermenge, in eine kleinere Zusammenstellung oder in ein neues Format. Die Transformationen können aber auch die Veränderung der Erhaltungsmetadaten in der AIU vor der Auslieferung umfassen.

Das OAIS für digitale Filme könnte es einem Benutzer zum Beispiel erlauben, einen digitalen Film als VHS-Band, als DVD oder als online gelieferte MPEG-Datei zu bestellen. Jede dieser Lieferungen würde eine Formatänderung erfordern und (theoretisch) eine Aktualisierung der Erhaltungsmetadaten im AIP, um korrekte Erhaltungsmetadaten für das DIP zu erzeugen.

4.2.2.7 Archivinformationskollektionen (AIC)

Die AIU und ihre zugehörigen Archiveinheits-Beschreibung liefern die Information, die ein Endnutzer braucht, um AIUs zu finden und zu bestellen. Es kann für einen Endnutzer aber nicht möglich sein, die Millionen von Archiveinheits-Beschreibungen durchsuchen, die in einem großen Archiv vorhanden sind. Um dieses Problem geht es hier.

Die Inhaltsinformation einer AIC setzt sich zusammen aus kompletten AIPs, die alle wieder ihre eigene Inhaltsinformation, Erhaltungsmetadaten und zugehörigen Verpackungsinformation und Paketbeschreibungen haben. Diese AIPs werden dann nach Kriterien, die der Archivar festlegt, zu einer Archivinformationskollektion (AIC) verbunden. Im Allgemeinen basieren AICs auf AIUs, die gemeinsame Themen oder Entstehungszusammenhänge sowie einen gemeinsamen Satz an Begleitbeschreibungen haben. Man kann sich vorstellen, dass alle OAIS im Minimum eine AIC haben, die alle AIPs des OAIS enthält.

Das OAIS für digitale Filme zum Beispiel kann AICs für die verschiedenen Film-Genres haben, wie Krimi, Science Fiction oder Horror. Zusätzlich kann das Archiv AICs haben, die nach anderen Eigenschaften organisiert sind wie Regisseur oder Hauptdarsteller.

Ein logisches Modell einer AIC wird in Abbildung 4-23 gezeigt. Wie in den vorherigen Unterabschnitten auch sind alle „Enthält“-Beziehungen logische Beinhaltungen, die entweder physisch sind oder durch den Verweis auf ein anderes Objekt im Magazin realisiert werden. Beispielsweise kann die Inhaltsinformation einer AIC entweder erzeugt werden, indem Kollektionen der enthaltenen AIPs physisch hergestellt werden, oder indem man auf die enthaltenen AIPs verweist. Ein einzelnes AIP kann zu beliebig vielen AICs gehören.

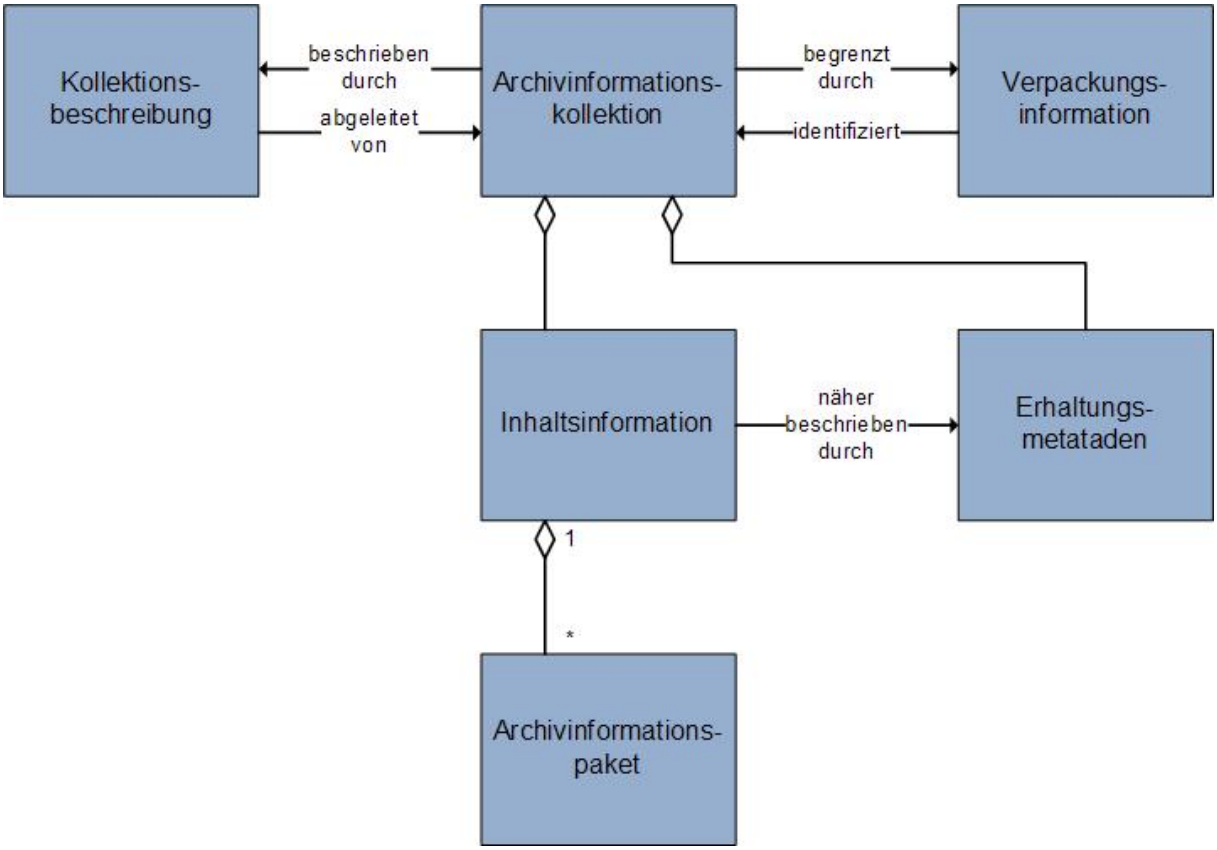

Abbildung 4-23: Logische Sicht einer Archivinformationskollektion

Zum Beispiel könnte eine Technik entwickelt werden, die Muster in digitalen Filmen erkennt, und das OAIS für digitale Filme könnte einen Dienst anbieten, der seine Bestände durchsucht nach großen Gebäuden wie den Pyramiden oder der New Yorker Skyline. Diese Art von Dienst ist sehr rechenintensiv und betrifft eine potenziell sehr große Anzahl von AIUs, die vom ‚Archivspeicher' zum ‚Zugriff' übermittelt werden und dann den entsprechenden Prozess zur Analyse der Inhaltsinformation jeder dieser AIUs durchlaufen müssen. Wenn die Ergebnisse von allgemeinem Nutzen sind, könnte der Archivar die Ergebnisse dieser „inhaltsbezogenen Anfrage“ zusammenfassen zu einer Begleitbeschreibung einer neuen AIC, die Filme mit großen Gebäuden enthält. Diese Technik wird häufig als „Data Mining" bezeichnet.

Wie Abbildung 4-23 zeigt, ist eine wichtige Eigenschaft der AIC die Tatsache, dass eine AIC ein komplettes AIP ist, das Erhaltungsmetadaten enthält. Die Erhaltungsmetadaten liefern weitere Information über die AIC wie Provenienzinformation zu Zeitpunkt und Gründen der Erstellung, Kontextinformation zu verwandten AICs, das erwünschte Maß an Sicherheits- bzw. Persistenz-Information und Information über Zugriffsrechte. Diese Information kommt zu den Erhaltungsmetadaten der enthaltenen AIPs hinzu. Sie sind oft notwendig, um einem Endnutzer Vertrauen in die Zuverlässigkeit einer AIC zu geben. Im oben erwähnten Beispiel beruht die Nützlichkeit einer AIC von Filmen, in denen große Gebäude vorkommen, bis zu

einem gewissen Grad auf dem verwendeten Algorithmus sowie auf die Provenienz-information, wann diese AIC entstanden oder zuletzt aktualisiert worden ist.

4.2.2.8 Kollektionsbeschreibungen

Die **Kollektionsbeschreibung** ist ein Untertypus der Paketbeschreibung. Sie hat zusätzliche Strukturen, um besser mit der komplexen Inhaltsinformation einer AIC umgehen zu können. Die Kollektionsbeschreibung, deren Modell Abbildung 4-24 zeigt, enthält die Klassen von Information, die in der Archiveinheits-Beschreibung enthalten sind.

Es gibt zwei Arten von Begleitbeschreibungen in einer Kollektionsbeschreibung:
- Eine Übersichtsbeschreibung, die die Kollektion als Ganze beschreibt.
- Keine oder mehrere Teilbeschreibungen, die jeden einzelnen Bestandteil der Kollektion separat beschreiben.

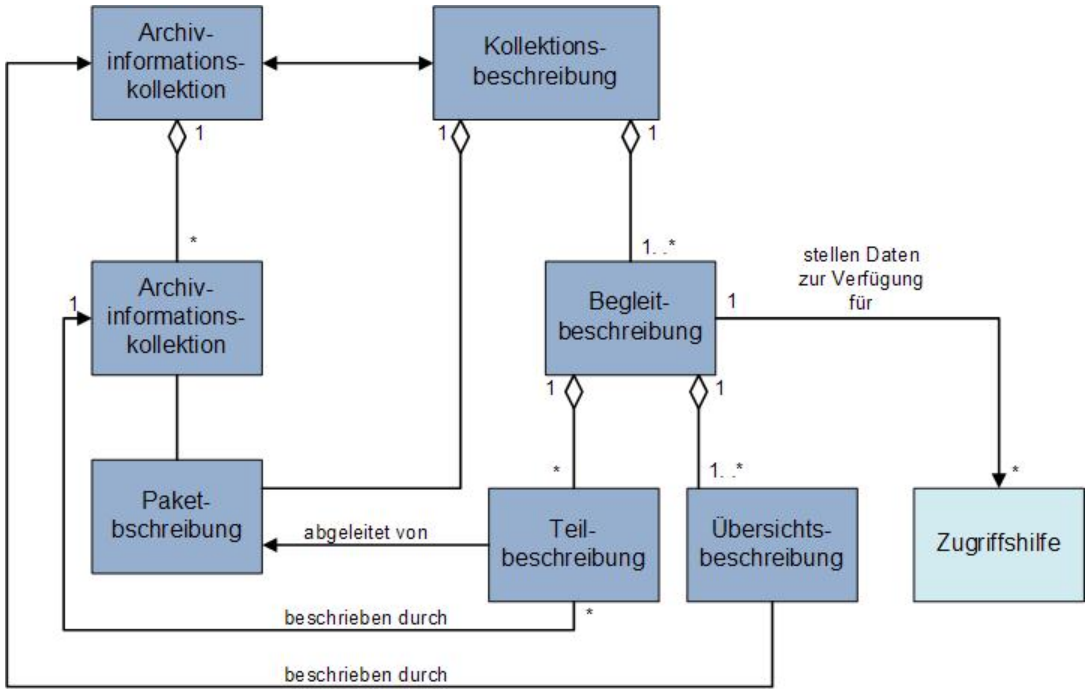

Abbildung 4-24: Kollektionsbeschreibungen

Die erforderliche Begleitbeschreibung in einer Kollektionsbeschreibung liefert Information für Bestellhilfen, die dem Benutzer den Zugriff auf den ganzen Satz von Inhaltsinformationen der entsprechenden AIC sowie auf deren Erhaltungsmetadaten gewährleisten, aber nicht notwendigerweise zu den einzelnen AIPs, die in der AIC enthalten sind. Die Kollektionsbeschreibung kann die Paketbeschreibungen der AIPs beinhalten, die in der AIC enthalten sind. Diese „Enthält"-Beziehung ist insofern eine logische, als dass die AIC entweder die Paketbeschreibungen von zugehörigen Informationspaketen direkt beinhaltet, oder – was häufiger der Fall ist – Verweise auf die Paketbeschreibungen der zugehörigen Informationspakete verwendet. Diese Liste von Paketbeschreibungen für in einer AIC enthaltene AIPs könnte Zugriffshilfen mit einer Methode versorgen, um einzelne AIPs einer AIC zu suchen und zu bestellen.

Sie lässt auch alternative Konzepte für die Implementierung von Findmitteln zu, die es dem Endnutzer ermöglichen, in einer AIC enthaltene AIPs zu finden. Die Begleitbeschreibungen, die Daten für diese Findmittel bereitstellen, können entweder zentral implementiert werden, wobei man eine Begleitbeschreibung in der Kollektionsbeschreibung sucht, oder dezentral,

wobei man die Begleitbeschreibung in den einzelnen Paketbeschreibungen aller enthaltenen Archivinformationspakete sucht.

Ein weiterer wichtiger Nutzen der Kollektionsbeschreibungen liegt in der Möglichkeit, neue **Zugriffskollektionen** zu definieren. Eine Zugriffskollektion kann auf neuen Resultaten von „Data Mining“ basieren, oder sie kann aktuelle Phänomene oder Themengebiete widerspiegeln, die nicht von dauerhaftem Interesse sein werden. Beispiele für eine Zugriffskollektion in einem OAIS für digitale Filme könnte eine Kollektion neuer Filmzugänge sein oder die Kollektion der 20 populärsten Filme, die regelmäßig aktualisiert wird. Ein anderes Beispiel für eine Zugriffskollektion ist eine Kollektion, die auf den Ergebnissen eines Algorithmus zur Mustererkennung basiert, der noch nicht verifiziert worden ist.

Um eine Zugriffskollektion zu erzeugen, erzeugt ein Archiv eine Kollektionsbeschreibung ohne zugehörige AIC. Die Kollektionsbeschreibung kann eine individuell angepasste, begleitende Teilbeschreibung haben, die die neu ausgewerteten Beschreibungsdaten für jedes zugehörige AIP dokumentiert. Ein spezielles Findmittel kann diese neuen begleitenden Teilbeschreibungen zusammen mit den bestehenden Begleitbeschreibungen in der Paketbeschreibungsinformation eines jeden zugehörigen AIPs benutzen, um die für den Benutzer interessanten AIPs zu finden. Die Paketbeschreibungen der enthaltenen AIPs würden außerdem Daten liefern für ein Bestellsystem, mit dem Endnutzer die relevanten Informationspakete bestellen können.

Wenn ein OAIS beschließt, dass eine Zugriffskollektion wertvoll genug
ist, um langfristig erhalten zu werden, kann es die nötigen Inhaltsinformationen und Erhaltungsmetadaten in den ‚Archivspeicher’ einstellen, um so eine neue AIC zu schaffen.

Eine weitere wichtige Anwendung von Zugriffskollektionen ist das Konzept, Teile einer Kollektion zu finden, deren Übernahme in das Archiv erst geplant ist. In diesem Fall würde die Begleitbeschreibung ein Findmittel unterstützen und das Auffinden zukünftiger AIPs ermöglichen. Die Begleitbeschreibung für das Bestellsystem und/oder das Bereitstellungssystem würde die Information enthalten, dass dieses Produkt gegenwärtig nicht erhältlich ist, und es dem Benutzer ermöglichen, eine ereignisbasierte Bestellung aufzugeben, die ausgeführt wird, sobald das gewünschte AIP verfügbar ist.

4.2.3. Information der Datenverwaltung

Gegenwärtig werden Paketbeschreibungen in dauerhaften Speichern wie Datenbankmanagementsystemen aufbewahrt, um leicht und flexibel den Zugriff auf und die Aktualisierung der enthaltenen Begleitbeschreibungen zu ermöglichen. Zusätzlich zu den Paketbeschreibungen, die in den vorherigen Unterabschnitten diskutiert worden sind, kann alle Information, die für den Betrieb eines Archivs gebraucht wird, in Datenbanken als dauerhafte Datenklassen gespeichert werden. Abbildung 4-25 zeigt die verschiedenen Arten von “Datenverwaltungs-Informationen“ innerhalb eines OAIS. Die Information der Archivadministration stellt die gesamte Bandbreite der Information dar, die für den täglichen Betrieb des Archivs nötig sind. Diese Information beinhaltet:

- Policy-Information, welche Gebühren und Zugangsbeschränkungen beim Bestellen von archivierter Information bereitstellen.

- Information zum Status einer Anfrage, die den Verlauf jeder Anfrage eines Benutzers beim Archiv aufzeichnen. Der Statusabfrage-Prozess kann sehr kompliziert sein und viele Datenbankereignisse und -trigger beinhalten oder aber auch ganz einfach die Form einer Datei haben, die Bestellanfragen verfolgt.

- Sicherheitsinformationen, die Benutzernamen und Passwörter oder andere Mechanismen beinhalten, die gebraucht werden, um die Identität und die Rechte der Benutzer zu authentifizieren.

- Ereignisbasierte Bestellungs-Informationen, die Information zur Unterstützung sich wiederholender oder zukünftiger Anfragen bereitstellen.

- Statistische Information, die von der Archivadministration und vom Management gebraucht werden, um zukünftige Policies festzulegen und die Arbeitsleistung für einen effektiveren Archivbetrieb zu optimieren. Beispiele für diese Statistiken umfassen die Häufigkeit der Bestellung eines AIPs über einen gewissen Zeitraum und die durchschnittliche Zeit zwischen dem Eingang einer Bestellung und der Ausgabe des bestellten Bestands.

- Information über den Verlauf des Erhaltungsprozesses, die die Migration von AIPs nachverfolgt inklusive des Ersetzens von Datenträgern und der Transformation von AIPs.

- Kundenprofil-Information, die es dem Archiv ermöglichen, Information wie Namen oder Adresse eines Benutzers nachzupflegen. So kann vermieden werden, dass der Benutzer diese Angaben bei jeder Bestellung von Neuem eintragen muss.

- Buchhaltungsinformation, die Daten enthalten, die für den Betrieb eines Archivs als wirtschaftliches Unternehmen notwendig sind. Die Buchhaltungsdaten beinhalten Lohnlisten sowie Angaben über Ausgaben und Einnahmen.

Diese Klassen sind eher als Beispiel denn als vollständige Liste aller Daten gemeint, die für die Administration eines Archivs gebraucht werden. Es handelt sich um konzeptionelle Klassen, und einzelne OAIS-Implementierungen können sich erheblich unterscheiden. Zum Beispiel kann ein einzelnes OAIS beschließen, die kundenbezogene Information wie Sicherheitsinformation und Kundenprofile in einer einzigen Datenbank zu kombinieren.

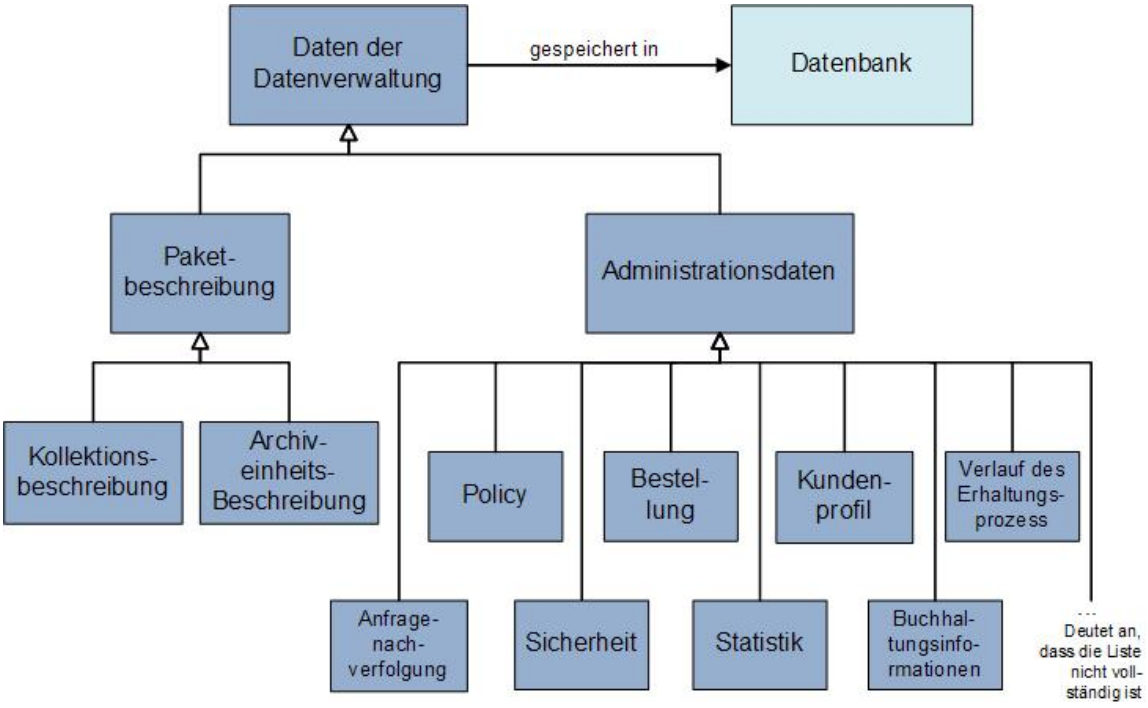

Abbildung 4-25: Datenverwaltungs-Information

4.3. Transformationen von Informationspaketen

In den vorherigen Teilen dieses Abschnitts ging es um die funktionale Architektur eines OAIS und um eine Informations-Architektur zur Darstellung der Informationspakete und ihrer zugehörigen Paketbeschreibungen und Verpackungsinformationen. Dieser Unterabschnitt betrachtet die logische und physische Transformation des Informationspakets und der damit verbundenen Objekte während ihres Lebenszyklus vom Produzenten zum OAIS und vom OAIS zum Endnutzer.

Abbildung 4-26 zeigt in einem Datenflussdiagramm die wesentlichen Datenflüsse im Überblick, die Anteil am OAIS-Betrieb haben. Dieser Datenfluss berücksichtigt keine Verwaltungsvorgänge wie den Rechnungsverkehr.

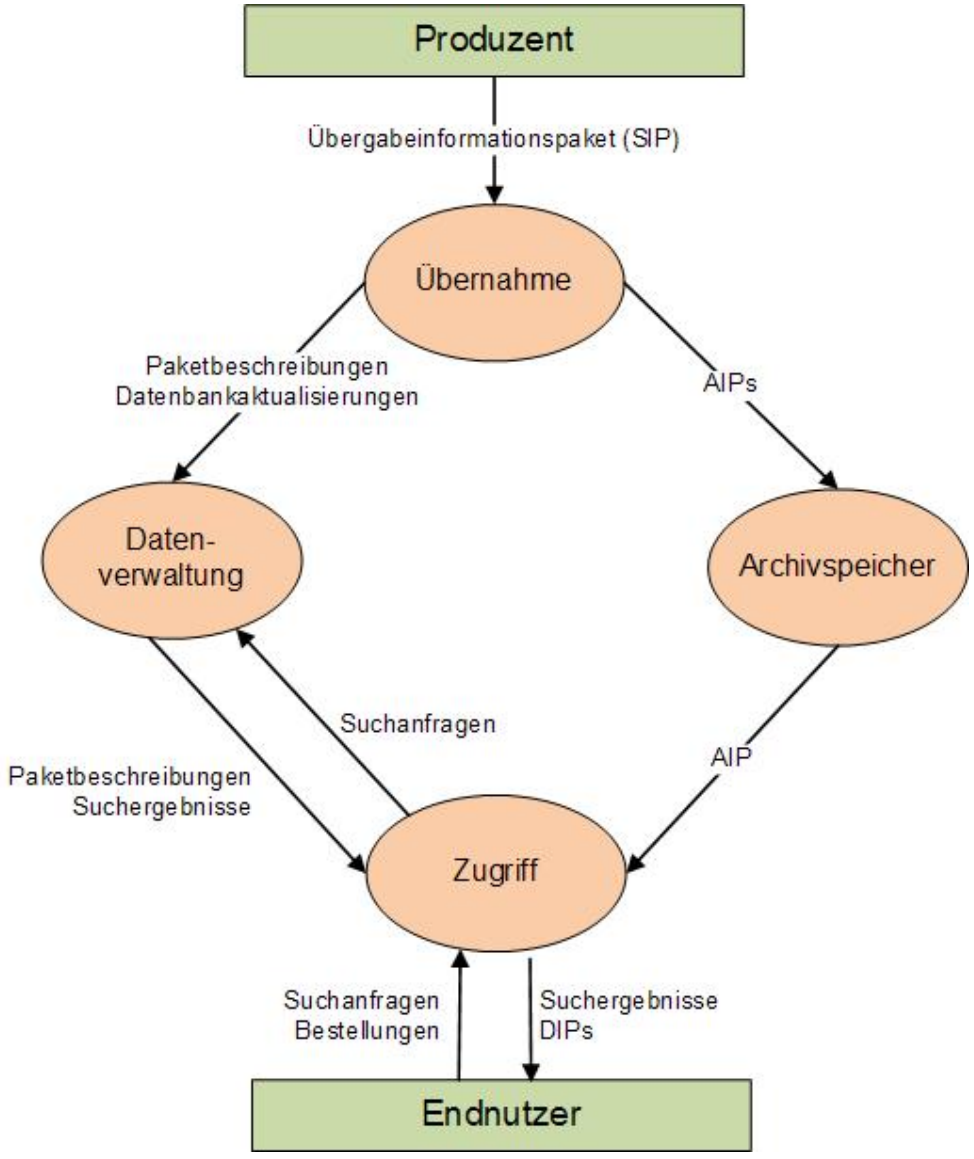

Abbildung 4-26: Datenflüsse in einem OAIS im Überblick

4.3.1. Datentransformation in der Einheit „Produzent“

Die Daten in der Einheit des Datenproduzenten sind nichtöffentlich und können in jedem vom Produzenten gewünschten Format vorliegen. Wenn allerdings die Entscheidung getroffen wurde, die Daten in einem OAIS zu speichern, trifft sich der Produzent, der für die Daten verantwortlich ist, mit den Archivaren, um eine Übergabevereinbarung auszuhandeln, wie in 2.3.2 diskutiert. Diese Vereinbarung definiert Information wie den Inhalt, das Format und die geplanten Eingangszeiten des Übergabeinformationspakets (SIP). Das SIP ist ein Informationspaket, das vom Produzenten an das OAIS übergeben wird. Das SIP besteht aus Inhaltsinformation und den notwendigen Daten, die sicher stellen, dass diese Daten vom Archiv auch gepflegt und von Endnutzern interpretiert und benutzt werden können, die sie in der Zukunft aus dem OAIS abrufen.

Diese SIPs werden regelmäßig in einer Datenübergabesitzung an das OAIS überführt. Die Anzahl der Datenübergabesitzungen zwischen einem OAIS und einem Produzenten kann von einer einzigen Sitzung für den Transfer eines abgeschlossenen Datenprodukts bis hin zu mehreren Sitzungen am Tag bei einem aktiven OAIS reichen, das Daten von Experimenten speichert, die sich noch in Bearbeitung befinden. Die Datenübergabesitzung kann logisch als Gruppen von Inhaltsdatenobjekten und Beschreibungs-Objekten betrachtet werden, wobei die Beschreibung physisch in den digitalen Objekten (d.h. selbstbeschreibende Objekte) enthalten oder in viele verschiedene beschreibende Einheiten unterteilt sein kann. Zusätzlich zu der logischen Sicht auf die Daten (das SIP) muss die Spezifizierung einer Datenlieferungssitzung auch die Zuordnung der Objekte zu den Datenträgern beinhalten, auf denen sie abgeliefert werden. Diese Zuordnung beinhaltet die Kodierung der Objekte und ihrer Beschreibung sowie die Zuweisung von logischen Objekten zu Dateien.

4.3.2. Datentransformationen in dem Funktionsbereich ‚Übernahme‘

Sobald sich das SIP im OAIS befindet, können sich dessen Form und Inhalt ändern. Vom OAIS wird nicht immer verlangt, die ihm übergebene Information in exakt demselben Format wie im SIP zu erhalten. Es kann sogar sein, dass die Erhaltung der Originalinformation in exakt der übergebenen Form nicht wünschenswert ist. Beispielsweise kann der Datenträger, auf dem übergebene Bilder gespeichert sind, technisch obsolet werden, und die Bilder müssen auf ein zeitgemäßeres Medium kopiert werden. Zudem werden einige Arten von Information, wie beispielsweise der eindeutige Identifikator, der zur Lokalisierung des Informationspakets innerhalb des OAIS dient, dem Produzenten nicht zugänglich sein und müssen während des Übernahmeprozesses in das OAIS eingegeben werden.

Die Zuordnung zwischen SIPs und AIPs ist nicht eins zu eins. Hier sind einige Beispiele:

- Ein SIP – Ein AIP: Eine Regierungsbehörde ist bereit, ihre elektronischen Akten des vorausgegangenen Steuerjahrs zu archivieren. Alle Akten des Jahres befinden sich auf Magnetbändern, die als ein SIP übergeben werden. Das Archiv speichert die Bänder zusammen als ein einziges AIP.

- Mehrere SIPs – Ein AIP: Ein Satellitensensor beobachtet über einen Zeitraum von einem Jahr die Erde. Jede Woche werden die neuesten Daten als SIP in das Archiv übertragen. Das Archiv hat ein einziges AIP, das alle Beobachtungen des Satelliten des ganzen Jahres enthält. In der ‚Übernahme‘ wird die Inhaltsinformation von jedem wöchentlichen SIP zu einer oder mehreren festgelegten Dateien im dauerhaften Übernahme-Speicher zusammengeführt. Die Erhaltungsmetadaten für das AIP werden gesendet, nachdem die letzten Sensordaten des Jahres angekommen sind. Nachdem alle wöchentlichen SIPs und das SIP mit den Erhaltungsmetadaten angekommen sind, generiert die ‚Übernahme‘ das AIP.

- Ein SIP – Mehrere AIPs: Eine Firma überträgt Finanzakten als ein SIP zu einem Archiv. Das Archiv entschließt sich, diese Information in zwei AIPs zu speichern: eines, das öffentliche Information enthält und eines, das sensible Information enthält. So ist es für das Archiv einfacher, den Zugriff zu der Information zu verwalten.

- Mehrere SIPs – Mehrere AIPs: Eine Öl- und Gasgesellschaft sammelt Information über ihre Quellen. Jedes Jahr übergibt sie SIPs mit aller Information einer Quelle an ein Archiv. Das Archiv führt ein AIP für jedes Öl- oder Gasfeld und löst die Information über jede Quelle heraus, um sie entsprechend ihrer geographischen Koordinaten dem richtigen AIP hinzuzufügen.

Der Übernahmeprozess wandelt die in der Datenübergabesitzung empfangenen SIPs in einen Satz von AIPs und Paketdeskriptoren um, die von den Funktionseinheiten ‚Archivspeicher' und ‚Datenverwaltung' gespeichert und angenommen werden können. Die Komplexität dieses Übernahmeprozesses kann von OAIS zu OAIS bzw. von Produzent zu Produzent innerhalb eines OAIS stark variieren. Die einfachste Form des Prozesses beinhaltet die Extraktion von Inhaltsinformation, Erhaltungsmetadaten und Paketdeskriptoren vom Transfermedium des Produzenten und deren Einreihung für die Speicherung in den Funktionseinheiten ‚Archivspeicher' und ‚Datenverwaltung'. In komplexeren Fällen müssen die Erhaltungsmetadaten und die Paketdeskriptoren aus der Inhaltsinformation extrahiert werden oder während des Übernahmeprozesses von den Mitarbeitenden des OAIS eingegeben werden; die Kodierung der Informationsobjekte oder ihre Zuordnung zu Dateien muss vielleicht verändert werden. Im extremsten Fall kann die Granularität der Inhaltsinformation verändert werden, und das OAIS muss neue Erhaltungsmetadaten und Paketdeskriptoren generieren, die die neu erzeugten Informationsobjekte wiedergeben. Wenn viele SIPs zur Erzeugung eines AIPs benötigt werden, wird der Funktionsbereich ‚Übernahme' einen Zwischenspeicher für die SIPs zur Verfügung stellen, bis alle SIPs eingetroffen sind, die für das AIP benötigt werden.

Zudem wird die Funktionseinheit ‚Übernahme' eintreffende Informationsobjekte klassifizieren, bestimmen, in welche bereits bestehende Kollektion oder Kollektionen jedes Objekt gehört, und Benachrichtigungen erstellen, um die entsprechenden Kollektionsbeschreibungen zu aktualisieren, nachdem die AIPs im ‚Archivspeicher' gespeichert wurden. Das OAIS und externe Organisationen können zusätzliche Begleitbeschreibungen und Findmittel bereitstellen, die alternative Zugangswege zu den gewünschten Informationsobjekten erlauben. Forscher werden neue und grundlegend unterschiedliche Zugriffsmuster zu den Informationsobjekten entwickeln. Es ist wichtig, dass die ‚Übernahme' und die internen Datenmodelle eines OAIS so flexibel sind, dass sie diese neuen Beschreibungen einbauen können und der allgemeine Nutzerkreis so von den Forschungsbemühungen profitiert. Ein gutes Beispiel für diese Art von neuer Begleitbeschreibung wäre eine Phänomenologie-Datenbank in der Erdbeobachtung, die es dem Benutzer erlaubt, mit einer einzigen Suchanfrage Daten von vielen Instrumenten für ein bestimmtes Ereignis, wie z.B. einen Wirbelsturm oder einen Vulkanausbruch, zu erhalten. Es ist wichtig zu bedenken, dass solche Findmittel obsolet werden können, sofern die von ihnen benötigten Daten nicht als Teil der AIPs gespeichert sind, auf die sie zugreifen.

Es wird erwartet, dass die Funktionseinheit ‚Übernahme' die Aktualisierungen zwischen ‚Datenverwaltung' und ‚Archivspeicher' koordinieren sowie eine angemessene Koordination und Fehlerbehebung bieten wird. Das AIP sollte zuerst im ‚Archivspeicher' gespeichert werden. Die Bestätigung dieser Operation beinhaltet eine eindeutige Identifikation, die zur Bereitstellung dieses AIPs aus dem Speicher dient. Dieser Identifikator sollte in die Paketbeschreibung übergehen, bevor die Kollektionsbeschreibung der ‚Datenverwaltung' hinzugefügt wird.

4.3.3. Datentransformationen in den Funktionsbereichen ‚Archivspeicher' und ‚Datenverwaltung'

Die Funktionseinheit ‚Archivspeicher' übernimmt die AIPs, die während des Übernahmeprozesses erzeugt werden, und überführt sie in den dauerhaften Archivbestand. Die Funktionseinheit ‚Datenverwaltung' übernimmt die Paketbeschreibungen, die von der ‚Übernahme' produziert wurden, und erweitert die existierenden Kollektionsbeschreibungen, um ihre Inhalte einzubinden. Das logische Modell der übernommenen Daten sollte bereits auf das logische Modell im Archivbestand abgebildet sein. Somit ist die wichtigste Transformation, die in diesem Schritt auftritt, die Abbildung der Zugangssitzung vom physischen Übernahme-Datenmodell (die tendenziell auf einem temporären Speicher sein wird) zum dauerhaften Speicher des OAIS, der von einem Datenbankmanagementsystem (DBMS) bis hin zum Hierarchischen Dateiverwaltungssystem (HFMS), oder einer Mischung beider Systeme reichen kann.[18]

Die interne Sicht des OAIS ist die dauerhafte Repräsentation der archivierten Daten. Deshalb müssen alle Kodierungen und Zuordnungen gut dokumentiert und verstanden sein. Der Transfer der Übernahme-Objekte findet häufig mit einem Softwareprozess wie einem HFMS-Treiber oder einem Datenbankmanagementsystem statt. In diesem Fall ist das OAIS dafür verantwortlich, eine aktive Kopie der Software oder eine sorgfältige Dokumentation der internen Formate zu unterhalten, so dass die Daten in der Zukunft ohne Informationsverlust in andere Systeme übertragen werden können.

4.3.4. Datenfluss und -transformationen in der Funktionsbereich ‚Zugriff'

Wenn ein Endnutzer die Daten in einem OAIS nutzen möchte, kann er ein Findmittel verwenden, um relevante Information zu finden. Findmittel präsentieren den Endnutzern die logische Sicht auf die Bestände des OAIS, so dass die Endnutzer entscheiden können, welche AIPs sie haben wollen. Die Zugriffs-Sicht ist mindestens die logische Sicht auf hoher Ebene auf die Kollektionsbeschreibungen, wie in 4.2.2.8. beschrieben. Es kann sein, dass das OAIS erhebliche Zeit und Mühe darauf verwenden muss, Begleitbeschreibungen und Findmittel zu entwickeln, wie z.B. Kataloge, die dem Endnutzer beim Auffinden von relevanten AIPs oder AICs helfen. Ein Endnutzer wird eine Recherchesitzung in der Funktionseinheit ‚Zugriff' beginnen. Während dieser Recherchesitzung wird der Endnutzer die Findmittel des OAIS benutzen, um möglicherweise relevante Bestände zu finden und zu untersuchen. Dieser Suchprozess ist in der Regel iterativ, da zuerst grobe Kriterien bestimmt werden und im Anschluss daran die Kriterien auf Basis der vorangegangenen Suchergebnisse verfeinert werden. Wenn möglicherweise relevante Objekte gefunden worden sind, können detailliertere Findmittel verwendet werden, z.B. eine Bildbetrachtungssoftware oder Animationen, um das Suchergebnis weiter zu verfeinern.

Sobald der Endnutzer diejenigen Bestände des OAIS gefunden hat, die er bekommen will, benutzt er ein vom OAIS bereitgestelltes Bestellsystem, um eine Bestellung aufzugeben und die gewünschten Daten zu bekommen. Der Endnutzer erstellt eine logische Sicht der gewünschten AIPs und der dazugehörenden Paketbeschreibungen, die in dem Auslieferungsinformationspaket enthalten sein sollen, und spezifiziert die physischen Details der

[18] Die Passage ist im Original nach Erachten der Übersetzer verstümmelt. Das Original lautet: "Thus the major transformation that occurs in this step is the mapping of the acquisition session from the ingest physical data model. This will tend to be on temporary storage, to the permanent storage of the OAIS, which could range from a Database Management Systems (DBMS) to a Hierarchical File Management Systems (HFMS), or any mixture of the above." Die Übersetzer gehen davon aus (und haben entsprechend übersetzt), dass gemeint war: "Thus the major transformation that occurs in this step is the mapping of the acquisition session from the ingest physical data model (This will tend to be on temporary storage), to the permanent storage of the OAIS, which could range from a Database Management Systems (DBMS) to a Hierarchical File Management Systems (HFMS), or any mixture of the above."

Datenauslieferungssitzung, wie z.B. die Art des Datenträgers und das Objektformat. Es kann sein, dass dieser Prozess keine sichtbare Interaktion zwischen dem Endnutzer und dem OAIS beinhaltet, wenn entsprechende Standardeinstellungen existieren. Diese Bestellung kann außerdem jede Art von Transformation umfassen, die der Endnutzer für die Erstellung von DIPs aus den AIPs wünscht.

Der Funktionsbereich ‚Zugriff' zeichnet dann die Bestellvereinbarung in dem Funktionsbereich ‚Datenverwaltung' auf. Wenn die Bedingungen zur Ausführung einer aufgezeichneten Bestellvereinbarung erfüllt sind, koordiniert der Funktionsbereich ‚Zugriff' die Rückmeldung. (Für viele Bestellvereinbarungen sind die Bedingungen unmittelbar erfüllt; andernfalls benachrichtigt die ‚Administration' den ‚Zugriff', sobald die Bedingungen erfüllt sind). Der ‚Zugriff' kontaktiert die Funktionsbereiche ‚Speicher' und ‚Datenverwaltung' und fordert die AIPs und dazugehörigen Paketbeschreibungen an, die zum Bestücken der vom Endnutzer gewünschten DIPs benötigt werden. Die Funktionsbereiche ‚Speicher' und ‚Datenverwaltung' erzeugen im Zwischenspeicher Kopien der angeforderten Objekte.

Der ‚Zugriff' transformiert diesen Satz von AIPs und die dazugehörenden Paketbeschreibung anschließend in einen Satz von DIPs und speichert diese DIPs auf physischen Datenträgern (entweder physisch oder per Datenübertragung), die dem Endnutzer in einer Datenauslieferungssitzung geliefert werden. Die Komplexität dieses Transformationsprozesses kann stark variieren, je nach dem, welche Art von Bearbeitungsservice das OAIS anbietet und welchen Service der Endnutzer in seiner Bestellung verlangt. Im einfachsten Fall beinhaltet das DIP Duplikate der AIPs und der dazugehörenden Paketbeschreibungen aus dem ‚Archivspeicher' und der ‚Datenverwaltung'. In komplexeren Fällen müssen die gewünschten Inhaltsinformationen aus den Informationsobjekten extrahiert werden oder in selbstbeschreibende Informationsobjekte eingefügt werden, wobei die Kodierung der Informationsobjekte und ihre Aufteilung auf physische Dateien eventuell geändert werden müssen. Im extremsten Fall, wenn das OAIS Subsetting-Dienste unterstützt, kann die Granularität der Informationsobjekte verändert werden, und der Auslieferungsprozess muss DIPs und entsprechende Paketbeschreibungen generieren, die diese neue Granularität wiedergeben. Die Zuordnung zwischen DIPs und AIPs ist 1:1, solange keine Transformationen verlangt werden. Allerdings kann die Benutzung von Subsetting-Diensten und anderen Verarbeitungsoptionen mehrere DIPs aus einem einzigen AIP erzeugen oder aber ein einziges DIP, das auf einer Kombination vieler AIPs basiert.

5 Perspektiven der Erhaltung

Dieser Abschnitt beschreibt verschiedene Verfahren, die angewendet werden oder angewendet werden könnten, um digitale Information und Zugriffsservices zu digitaler Information zu erhalten. Es stützt sich auf die Funktions- und Prozessmodelle, die in Unterabschnitt 4.2 beschrieben wurden, wendet sie auf die verschiedenen Verfahren an und erweitert die Terminologie, um wesentliche Aspekte der verschiedenen Verfahren zu unterscheiden. Unterabschnitt 5.1 behandelt die Erhaltung digitaler Information durch Migration auf andere Datenträger und in andere Formate. In Unterabschnitt 5.2 geht es um die Erhaltung von Zugriffsservices zu digitaler Information, während sich Technologie verändert und Software auf neue Systeme übertragen wird, sie „eingehüllt“ wird, um einheitliche Schnittstellen instand zu halten, oder emuliert wird, um Alt-Anwendungen weiter zu unterstützen. Außerdem werden einige zentrale Fragen der verschiedenen Herangehensweisen identifiziert.

5.1 Informationserhaltung

Das sich schnell ändernde Wesen der Computerindustrie und das flüchtige Wesen von elektronischen Speichermedien stehen im Widerspruch zum Hauptanliegen eines OAIS: Information langfristig zu erhalten. Unabhängig davon, wie gut ein OAIS seinen aktuellen Bestand pflegt, es wird schließlich gezwungen sein, einen Großteil des Bestands auf andere Datenträger zu migrieren (was die Veränderung von Bit-Sequenzen zur Folge haben kann oder auch nicht) und/oder in eine andere Hardware- oder Software-Umgebung zu migrieren, um ihn zugänglich zu halten. Heutige Speichermedien können für gewöhnlich nur wenige Jahrzehnte eingesetzt werden, bevor die Wahrscheinlichkeit eines unwiderruflichen Datenverlusts zu groß wird, um ihn zu ignorieren. Das rasante Tempo der technologischen Entwicklung macht ferner viele Systeme bereits nach einigen Jahren deutlich weniger kosteneffektiv. Zusätzlich zu Technologieveränderungen wird es Veränderungen im Grundwissen der vorgesehenen Zielgruppe geben, was die benötigten Repräsentationsinformationen beeinflusst.

Es gibt etliche grundlegende Ansätze zur Informationserhaltung. Beim ersten verbleibt das Inhaltsdatenobjekt in seiner ursprünglichen Form, und Zugriff und Nutzung werden dadurch erreicht, dass ausreichende Beschreibungen der digitalen Codierung mit Struktur- und semantischer Repräsentationsinformation zur Verfügung gestellt werden; in manchen Fällen sind die ursprünglichen Zugriffs- und Nutzungsmechanismen ausreichend, in welchem Fall Software-Emulation (unter Nutzung anderer Repräsentationsinformation) nützlich sein mag, obwohl dies die Tendenz hat, die Arten, auf die das Inhaltsdatenobjekts genutzt werden kann, einzuschränken. Alternativ kann das Objekt transformiert werden in eines, das mit zeitgemäßen Zugriffs- und Nutzungsmechanismen verarbeitet werden kann. Dies wird gemeinhin als Transformation bezeichnet, eine Art der Migration, welche unten diskutiert wird.

Die folgende Tabelle zeigt die verschiedenen Kombinationen der oben diskutierten Alternativen.

	Inhaltsdatenobjekt unverändert	Inhaltsdatenobjekt verändert
Zugriffsservice unverändert	Falls unter Nutzung der ursprünglichen Software ausführbar: Emulation. Falls der ursprüngliche Quellcode genutzt wird: Ausführbares Datenobjekt wiederherstellen.	Re-Implementierung des Zugriffsservice
Zugriffsservice verändert	Implementierung eines neuen Zugriffsservice auf Grundlage der Repräsentationsinformation, die das ursprüngliche Inhaltsdaten-objekt beschreibt	Implementierung eines neuen Zugriffsservice auf Grundlage der Repräsentations-information[19], die das neue Inhaltsdatenobjekt beschreibt.

Zur Erhaltung vorgesehene Informationseigenschaften können genutzt werden, um den Erfolg zu testen wo der Zugriffsservice unverändert ist, zum Beispiel die Darstellung eines Bildes oder einfachen Dokuments, oder die Ausführung einer Software.

Digitale Migration ist definiert als Transfer digitaler Information innerhalb des OAIS zum Zweck ihrer Erhaltung. Sie unterscheidet sich von allgemeinen Transfers in drei Punkten:

- der Fokus liegt auf der Erhaltung des gesamten Informationsgehalts, der zur Erhaltung vorgesehen ist;

- es wird angenommen, dass die neue archivische Erscheinung der Information die alte ersetzt; und

- dem OAIS obliegt die volle Kontrolle und Verantwortung über alle Aspekte des Transfers.

Dieser Unterabschnitt befasst sich mit der digitalen Migration von AIPs innerhalb eines OAIS.

5.1.1 Gründe für eine digitale Migration

Es gibt drei wesentliche Gründe, digitale Migration von AIPs innerhalb eines OAIS zu betreiben:

- Verbesserte Wirtschaftlichkeit: Das rasante Tempo der Hardware- und Softwareentwicklung sorgt für erheblich ansteigende Speicherkapazitäten und Transferbandbreiten bei gleichzeitig immer geringeren Kosten. Es bringt außerdem mit sich, dass einige Speichermedien lange vor ihrem physischen Verfallsdatum bereits veraltet sind. Zusätzlich können verbesserte AIP-Konstruktionen diese weniger abhängig von den ihnen zugrundeliegenden Speichermedien und Systemen machen, was die Migrationsbemühungen vereinfachen dürfte. Um kosteneffektiv zu bleiben, muss ein OAIS die Vorteile dieser Technologien nutzen. Abhängig von den jeweils eingesetzten Technologien muss das AIP möglicherweise auf neue Speichermedien übertra-

[19] Im Original steht hier nur „Representation“. Die Übersetzer gehen davon aus (und haben entsprechend übersetzt), dass die Stelle im Original fehlerhaft ist und hier analog zur linken Tabellenspalte auch „Representation Information“ stehen sollte.

gen werden, die vorher nicht unterstützt wurden. In diesem Fall müsste die Implementierung des AIPs überprüft und ggf. verändert werden, um Vorteil aus den neuen Technologien ziehen zu können.

- Neue Anforderungen aus dem Endnutzer-Service: Die Endnutzer eines OAIS erfahren ebenfalls die Vorzüge neuer Technologien und dementsprechend wachsen ihre Ansprüche an Art und Umfang der Serviceleistungen eines OAIS. Diese verstärkten Serviceleistungen erfordern möglicherweise neue Formen von DIPs, um bestimmte vorgesehene Zielgruppen zu bedienen, welche wiederum ein OAIS dazu antreiben neue Formen von AIPs vorzuhalten, um die Zahl der Konvertierungen bei der Ausgabe zu senken. Des Weiteren unterliegen AIPs für gewöhnlich Schwankungen in der Nachfrage und das OAIS wird eventuell unterschiedlich leistungsfähige Ebenen von Zugriffsmöglichkeiten bereitstellen müssen, um den Ansprüchen der Endnutzer über die Zeit hinweg genügen zu können. Das kann wahrscheinlich gewährleistet werden, indem einige AIPs auf unterschiedliche Speichermedien übertragen werden, die jeweils eine höhere oder niedrigere Leistungsstärke beim Zugriff bieten. Schließlich kann sich die vorgesehene Zielgruppe für ein AIP erweitern, was dazu führt, dass AIPs auf ihre Verstehbarkeit und Nutzbarkeit für die erweiterte Zielgruppe hin überprüft werden müssen. Das alles kann zur Migration eines AIPs innerhalb eines OAIS führen.

- Verfall der Speichermedien: Digitale Speichermedien werden mit der Zeit zunehmend unzuverlässig als sicherer Bewahrer von Bits. Sogar solche Speichermedien, die mit einer Fehlerkorrektur arbeiten, müssen schließlich ersetzt werden. Im Endeffekt erfordert der Verfall der Speichermedien die Übertragung der AIPs auf neuere Speichermedien.

Digitale Migrationen sind zeitaufwendig, kostenintensiv und setzen das OAIS einer hohen Wahrscheinlichkeit von Informationsverlust aus. Deshalb sollte ein OAIS sich unbedingt mit den Problemstellungen und Herangehensweisen der digitalen Migration auseinandersetzen.

5.1.2 Migrationskontext

Abbildung 5-1 führt diejenigen grundlegenden Konzepte aus den Funktions- und Informationsmodellen aus Abschnitt 4 zusammen, die sich auf Aspekte der Migration beziehen:

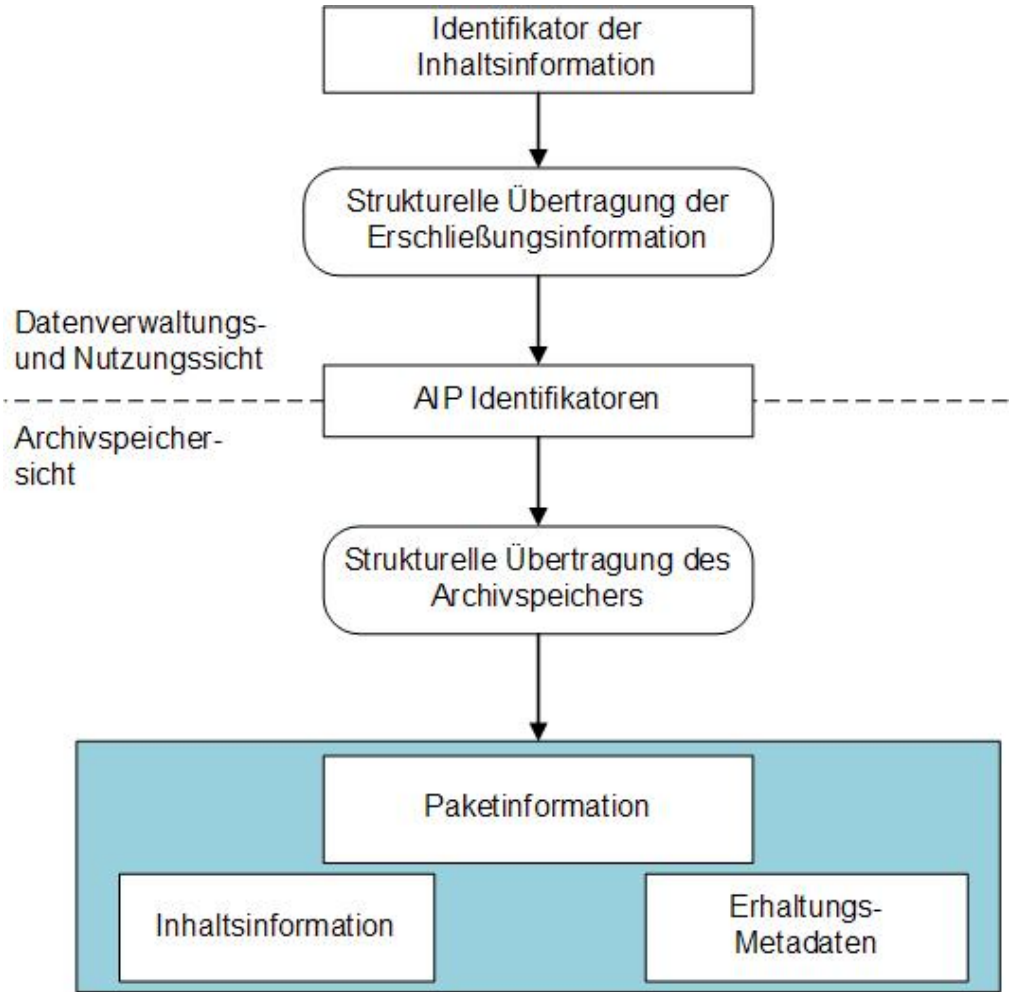

Abbildung 5-1: Konzeptionelle Sicht auf die Beziehungen zwischen Bezeichnungen und AIP-Bestandteilen

Die OAIS Endnutzer-Schnittstelle im ‚Zugriff' liefert eine oder mehrere Inhaltsinformations-IDs mit dazugehörigen Namensräumen, um bei der Identifizierung eines bestimmten Inhaltsinformationsobjekts zu helfen. Eine oder mehrere dieser Inhaltsinformations-IDs sind in der Referenzinformation der Erhaltungsmetadaten enthalten, die mit dem Inhaltsinformationsobjekt verknüpft sind. Die Erschließungsinformationen aus der ‚Datenverwaltung' wiederum übersetzen jede einzelne dieser IDs in dieselbe AIP-ID. Die Funktion ‚Zugriff' benutzt diese Information, um die AIP-ID abzurufen und übergibt sie an den ‚Archivspeicher', um das zugehörige AIP abzurufen.

Innerhalb des ‚Archivspeichers' wird die AIP-ID über die Infrastruktur des ‚Archivspeichers' auf den Speicherort der Verpackungsinformation abgebildet („mapping"). Die Verpackungsinformation ihrerseits grenzt die Inhaltsinformation und die Erhaltungsmetadaten logisch voneinander ab, identifiziert sie und verknüpft sie für die Erhaltung zu einer Einheit. Wenn zum Beispiel die Inhaltsinformation und die Erhaltungsmetadaten festgelegt wären sowohl als Inhalt mehrerer Dateien als auch als die Verknüpfungen mit den Dokumenten, welche die Repräsentationen dieser Dateien beschreiben sowie als die Dokumente selbst, dann wäre die Verpackungsinformation logisch definiert als Implementierung des Dateisystems, das die Bits der Dateien beinhaltet, als die Datenstruktur, welche die Verknüpfungen abbildet, als die

Information, die verwendet wird, um Inhaltsinformation und Erhaltungsmetadaten voneinander zu unterscheiden, und als eine kapselnde Datenstruktur, die Dateien und andere Datenstrukturen als Bestandteile des AIPs identifiziert. Die dazugehörige Archivspeicher-Mapping-Infrastruktur könnte in diesem Fall umgesetzt werden als Datenbank, welche die AIP-ID mit dem Speicherort der kapselnden Datenstruktur in Beziehung setzt.

Der Transfer eines jeden Teils der Inhaltsinformation, der Erhaltungsmetadaten oder der Verpackungsinformation auf einen gleichen oder neuen Datenträger mit dem Ziel, das er seinen Vorgänger im AIP ersetzt, wird als digitale Migration des AIPs verstanden. Eine Veränderung in der Archivspeicher-Mapping-Infrastruktur hingegen, die außerhalb des AIP-Konzepts steht, wird nicht als Migration des verknüpften AIPs verstanden, auch wenn solche Veränderungen sorgfältig kontrolliert werden sollten, um den Zugriff auf das AIP sicherzustellen.

Die Art der Implementierung der AIPs wird wesentlichen Einfluss haben sowohl auf den Grad der Automatisierung als auch auf die Wahrscheinlichkeit eines Informationsverlustes bei der Migration. Gut konstruierte AIPs können gleichzeitig den Grad der Automatisierung erhöhen und die Wahrscheinlichkeit eines Informationsverlustes verringern. Für ein besseres Verständnis des Einflusses dieser Faktoren auf die Migration von AIPs ist es sinnvoll, Migrationen in verschiedene Typen zu unterscheiden und dann einige Problemstellungen der verschiedenen Herangehensweisen zu betrachten.

5.1.3 Migrationstypen

Anhand der oben erwähnten Modelle und Konzepte lassen sich vier grundlegende Typen digitaler Migration unterscheiden, die nach dem steigenden Risiko eines möglichen Informationsverlustes angeordnet sind:

Verfahren, die die Bit-Sequenz nicht verändern:

- Auffrischung: Digitale Migration, bei der ein Datenträger, auf dem ein oder mehrere AIPs oder Teile von AIPs enthalten sind, durch einen Datenträger desselben Typs ersetzt wird, indem die Bits des Datenträgers kopiert werden, die AIPs enthalten und die benutzt werden, um den Datenträger zu verwalten und auf ihn zuzugreifen. Die bestehende Archivspeicher-Mapping-Infrastruktur ist ohne Änderungen weiterhin in der Lage, das AIP zu lokalisieren und darauf zuzugreifen.

- Replikation: Digitale Migration, bei der weder die Verpackungsinformation noch die Inhaltsinformation oder Erhaltungsmetadaten verändert werden. Die Bits zur Vermittlung dieser Informationsobjekte bleiben bei dem Transfer auf den gleichartigen oder neuen Datenträgertyp erhalten. Die Auffrischung ist also ebenfalls eine Replikation, eine Replikation könnte allerdings Veränderungen in der Archivspeicher-Mapping-Infrastruktur erforderlich machen.

Verfahren, die die Bit-Sequenz verändern:

- Umverpackung: Digitale Migration, bei der es einige Veränderungen in den Bits der Verpackungsinformation gibt.

- Transformation: Digitale Migration, bei der trotz einiger Veränderungen in den Bits der Inhaltsinformation oder Erhaltungsmetadaten versucht wird, den vollen Informationsgehalt zu erhalten.

Den geringsten Informationsverlust hat man bei der Auffrischung zu befürchten, weil keine Bits, die AIP-Information enthalten oder das Recherchieren von und den Zugriff auf AIPs unterstützen, verändert werden. Ein ebenso geringes Risiko des Informationsverlusts besteht im Falle der Replikation, weil keine Bits, die AIP-Information repräsentieren, verändert

werden. Wenn jedoch ein neuer Typ von Datenträger eingesetzt wird, werden einige Änderungen in der Archivspeicher-Mapping-Infrastruktur notwendig (s. Abbildung 5-1). Das Risiko besteht darin, dass während dieses Prozesses etwas schief gehen könnte und unvorhergesehene Änderungen an den Bits auftreten. Umverpackung kalkuliert von vorneherein einige Änderungen an den Bits ein, aber diese sind hauptsächlich beschränkt auf Information zur Abgrenzung der Inhaltsinformation und der Erhaltungsmetadaten und betreffen üblicherweise nicht die Information, die von den Inhaltsinformation und den Erhaltungsmetadaten transportiert werden. Es besteht das übliche Risiko, dass etwas schief gehen könnte und es gibt außerdem Fälle, in denen eine gegenseitige Beeinflussung von Verpackungsinformation und Inhaltsinformation oder Erhaltungsmetadaten nicht vermieden werden kann. Das bedeutet ein zusätzliches Risiko von Informationsverlust. Allerdings wird angenommen, dass das OAIS verifizieren wird, dass Auffrischungs-, Replikations- und Umverpackungs-Migrationen keine Information verloren haben. Transformation schließlich birgt das größte Risiko, weil Veränderungen an der Inhaltsinformation und an den Erhaltungsmetadaten vorgenommen werden.

Um besser verstehen zu können, welche Probleme mit diesen Migrationstypen einhergehen können, ist ein Blick auf mögliche Herangehensweisen zu ihrer Umsetzung nötig. Es wird sich herausstellen, dass einige Migrationen eine Mischform von Umverpackung und Transformation sind. Außerdem ist es wichtig, sich daran zu erinnern, dass das OAIS für jedes gegebene AIP zunächst die Zusammensetzung der Inhaltsinformation genau bestimmen muss, damit die Erhaltungsmetadaten entsprechend ermittelt werden können. Erst danach kann auch die Verpackungsinformation ermittelt werden. Desweiteren gibt es nicht die eine „korrekte“ Definition, was Inhaltsinformation ist. Das OAIS muss das für jedes AIP, das es konstruiert und aufbewahrt, festlegen. Alle diese Fragen werden in den folgenden Unterabschnitten anhand einer Reihe von praktischen Anwendungen und Migrationsszenarien näher diskutiert.

5.1.3.1 Auffrischung

Auffrischung wird eingesetzt, um einen Datenträger durch eine Kopie zu ersetzen, die so genau ist, dass Hard- und Software des Archivspeichers einfach weiterlaufen wie zuvor. Das folgende Szenario ist ein Beispiel für Auffrischung:
Die Zahl der korrigierbaren Bitfehler auf einer CD-ROM erreicht einen kritischen Punkt und es wird entschieden, sie durch eine genaue Kopie zu ersetzen. Sobald überprüft wurde, dass beide gleichwertig sind, ersetzt die neue CD-ROM die alte CD-ROM und die Auffrischung ist abgeschlossen. Alle AIP-Komponenten auf der CD-ROM sind unverändert.

5.1.3.2 Replikation

Bei einer Replikation gibt es keine Veränderungen in den Bits der Verpackungsinformation, der Inhaltsinformation oder der Erhaltungsmetadaten. Je nach Art der Umsetzung variiert der Aufwand, den man betreiben muss, um sicherzustellen, dass keines dieser Bits sich verändert hat. Das folgende Szenario ist ein Beispiel für Replikation:
Die Inhaltsinformation und die Erhaltungsmetadaten eines AIPs sind in eine standardisierte Verpackungsstruktur eingebunden und werden in Form einer einzigen Datei vorgehalten. Eine Replikation wird durch simples Kopieren der Bitabfolge der Datei in eine neue Datei auf dem gleichen oder einem anderen Datenträger erreicht. Je nach dem können Veränderungen in der Archivspeicher-Mapping-Infrastruktur notwendig werden, um die Datei weiterhin auffinden zu können, aber es wurden keine Veränderungen an der Verpackungsinformation, der Inhaltsinformation oder den Erhaltungsmetadaten vorgenommen. Mit diesem Typ von Verpackungsinformation lässt sich eine Replikation durch Migration auf neue Datenträger mit einem hohen Grad an Automatisierung und einem geringen Risiko des Informationsverlustes komfortabel durchführen.

5.1.3.3 Umverpackung

Bei einer Migration handelt es sich um eine Umverpackung, wenn sich die Verpackungsinformationen während des Transfers verändern. Die Verpackungsinformation spielt eine entscheidende Rolle, indem sie mindestens die Inhaltsinformation und Erhaltungsmetadaten voneinander abgrenzt und sie zueinander in Beziehung setzt. Setzen sich Inhaltsinformation und Erhaltungsmetadaten selbst wiederum aus verschiedenen Komponenten zusammen, muss die Verpackungsinformation auch diese voneinander abgrenzen und zueinander in Beziehung setzen. Das sind Umsetzungsentscheidungen, die das OAIS ausführlich berücksichtigen sollte. Das folgende Szenario ist ein Beispiel für Umverpackung:

> Alle Bits der Inhaltsinformation und der Erhaltungsmetadaten eines AIPs sind in Form von drei Dateien auf einer CD-ROM enthalten. Die Verpackungsinformation besteht aus den Bits, die gebraucht werden, um die Datei und die Dateiverzeichnisstruktur, die für den Zugriff auf diese drei Dateien sorgt, zu implementieren. Die Inhalte der drei Dateien werden in drei neue Dateien auf anderen Datenträgern übertragen, mit einem neuem Verzeichnis und einer neuen Implementierung. Selbst wenn alle Verzeichnis- und Dateinamen über den Transfer hinweg beibehalten wurden, hat eine Umverpackung stattgefunden, weil die Bits zur Repräsentation der Verpackungsinformation sich verändert haben.

5.1.3.4 Transformation

Digitale Migrationen, die einige Änderungen in der Inhaltsinformation oder den Erhaltungsmetadaten erfordern, werden als Transformation bezeichnet. Diese Änderungen werden an den Bits der Inhaltsinformation oder der Erhaltungsmetadaten vorgenommen und bringen entsprechende Änderungen in den zugehörigen Repräsentationsinformationen mit sich. In allen Fällen ist das Ziel maximale Informationserhaltung. Das neue AIP soll ein vollwertiger Ersatz für das alte AIP sein. Das neue AIP gilt als **transformiertes AIP** des vorherigen AIPs. Die erste Version des AIP wird als Original-AIP bezeichnet und kann zur Verifizierung der Informationserhaltung aufbewahrt werden.

Die Repräsentationsinformation spielt in der Transformation eine zentrale Rolle, und die Auswirkungen von Veränderungen an der Repräsentationsinformation können herangezogen werden, um die Transformationen zu kategorisieren. Das Repräsentationsinformationsobjekt kann aus einer Grundmenge von Einheiten, einer Menge von hieraus entstehenden Einheiten sowie Übersetzungs-Regeln gebildet werden, welche die entstehenden Einheiten und ihre Beziehungen in Bezug auf die Grundeinheiten genau beschreiben.
Unter Verwendung dieses Modells eines Repräsentationsinformationsobjekts lassen sich zwei Arten der Transformation definieren: **reversible Transformation** und **irreversible Transformation**.

Eine reversible Transformation geschieht, wenn die neue Repräsentation eine Menge (oder eine Teilmenge) von entstehenden Einheiten definiert, die den entstehenden Einheiten entsprechen, die von der originalen Repräsentation definiert wurden. Das bedeutet eine Eins-zu-eins-Abbildung zurück auf die originale Repräsentation und ihre Menge von Grundeinheiten. Ein Beispiel ist die Ersetzung einer Repräsentation, welche die ASCII-Codes für „A bis Z“ verwendet, durch eine Repräsentation, welche die UNICODE UTF-16-Codes für „A bis Z“ verwendet. Die Transformation ergibt die Ersetzung von 7-Bit-Zeichencodes durch 16-Bit-Zeichencodes in dem veränderten AIP-Objekt. Die Rück-Transformation könnte anschließend durch Ersetzung der UNICODE UTF-16-Codes für „A bis Z“ durch die ASCII-Codes für „A bis Z“ durchgeführt werden und das originale AIP wäre wiederhergestellt.

Eine irreversible Transformation tritt ein, wenn eine reversible Transformation nicht garantiert werden kann. Die Ersetzung eines IBM 7094 Gleitkommawerts durch einen IEEE Gleitkommawert ist zum Beispiel eine irreversible Transformation, weil die entstehenden Einheiten

dieser zwei Repräsentationen semantisch nicht gleichwertig sind. Eine ist exakter als die andere. Allerdings mögen sie – abhängig von ihrer Verwendung – ausreichend identisch sein, um praktisch austauschbar zu sein. In diesem Fall bewahrt eine irreversible Transformation erfolgreich den Informationsgehalt. Für komplexe Formate, in denen Bedeutungen und Beziehungen zwischen Gruppen wesentlich sind, dürfte es hingegen schwierig sein, zu begründen, dass eine irreversible Transformation die Inhaltsinformation ausreichend konserviert hat. Es ist nützlich, eine **transformationelle Informationseigenschaft** als eine Informationseigenschaft zu definieren, deren Erhaltung als notwendig, aber nicht ausreichend betrachtet wird, um zu beweisen, dass die irreversible Transformation den Informationsgehalt in ausreichendem Maß erhalten hat. Dies könnte wichtig sein, da es zum Nachweis der Authentizität beiträgt. Solchen Informationseigenschaften wird spezifische Repräsentationsinformation, einschließlich semantischer Information, zugeordnet werden müssen, um anzuzeigen, wie sie kodiert sind und was sie meinen. (Man beachte, dass der Begriff „Significant Property“, der in der Literatur unterschiedlich definiert wird, manchmal auf eine Art verwendet wird, die konsistent mit transformationeller Informationseigenschaft ist.) Beispiele für reversible und irreversible Transformationen werden in den folgenden Szenarien dargestellt.

Das folgende Szenario bezeichnet eine reversible Transformation, die auftritt, wenn eine verlustfreie Komprimierungs-Funktion in die Inhaltsinformation eines AIPs integriert wird.

> Alle Bits der Inhaltsinformation eines AIPs sind in Form von drei Dateien auf einer CD-ROM enthalten. Die Verpackungsinformation beinhaltet die Bits zur Implementierung der Datei- sowie der Verzeichnisstruktur, die den Zugriff auf diese drei Dateien bietet. Die Inhalte der drei Dateien werden auf eine neue CD-ROM übertragen und während dieses Prozesses werden sie mit einem verlustfreien Kompressions-Algorithmus komprimiert. Dieser Transfer ist eine Transformation, weil der Komprimierungs-Prozess die Inhaltsinformation modifiziert hat, und es handelt sich dabei um eine reversible Transformation, weil es einen Dekomprimierungs-Algorithmus gibt, der die originalen Bits des Dateiinhalts zurückbringt. Die relevanten Komponenten der Repräsentationsinformation der Original-Inhaltsinformation müssen aktualisiert werden, um den Dekomprimierungs-Algorithmus aufzunehmen, ebenso wie die Erhaltungsmetadaten aktualisiert werden müssen, um dieses neue transformierte AIP abzubilden.

Das folgende Szenario bezeichnet eine irreversible Transformation, die auftreten kann, wenn Inhaltsinformation in ein neues Format migriert werden, das ein differenzierteres Datenmodell darstellen kann als das Ausgangsformat.

> Alle Bits der Inhaltsinformation eines AIPs sind in Form von drei Dateien auf einer CD-ROM enthalten. Die Verpackungsinformation beinhaltet die Bits zur Implementierung der Datei- sowie der Verzeichnisstruktur, die den Zugriff auf diese drei Dateien bietet. Die Inhalte der drei Dateien werden auf eine neue CD-ROM übertragen und während dieses Prozesses wird die dritte Datei verändert, weil es nicht mehr ohne weiteres verfügbare Werkzeuge gibt, um die Datei in ihrer derzeitigen Form effektiv zu nutzen. Das neue Format ist zwar allgemein gebräuchlich, verwendet aber ein anderes Datenmodell als das Originalformat und es gibt verschiedene Möglichkeiten, die Information in das neue Format zu übersetzen. Diese Übersetzung muss sehr vorsichtig geschehen, um sicherzustellen, dass keine für die vorgesehene Zielgruppe signifikante Information verloren gehen. Zum Beispiel könnte eine transformationelle Informationseigenschaft für wissenschaftliche Daten die Werte von identifizierten Datenelementen zu einem festgelegten Genauigkeitsgrad sein; wenn die Inhaltsinformation ein Dokument wäre, könnte das Seiten-Layout eine transformationelle Informationseigenschaft sein.
>
> Diese Übersetzung muss in den Erhaltungsmetadaten enthalten sein und natürlich wird die Repräsentationsinformation, die das neue Format beschreibt, diejenige ersetzen, die das alte Format beschrieben hat. Das Ergebnis ist ein neues transformiertes AIP. Es handelt sich hier um eine Transformation, die außerdem eine irrever-

sible Transformation ist, wenn es keinen Algorithmus gibt, der die Original-Datei aus der neuen Datei reproduzieren kann.

Das folgende Szenario bezeichnet eine reversible Transformation, die Umverpackung einbezieht. Dieser Fall tritt ein, wenn in die Inhaltsinformation ein Dateiname eingebettet ist, der auf eine ihrer Komponenten verweist, und wenn die Inhaltsinformation mit neuen Dateinamen auf einen neuen Datenträger übertragen wird.

> Die Inhaltsinformation für ein AIP ist in Form von drei Dateien auf einer CD-ROM enthalten. Die erste Datei enthält einen internen Namen, der die erste Datei mit der dritten Datei verknüpft und die Beziehung zwischen den beiden Dateien näher beschreibt. Die Verpackungsinformation enthält das Dateiverzeichnis und die Dateistruktur zur Identifizierung der drei Dateien. Während der Migration auf einen neuen Datenträger werden diese drei Dateien umbenannt und in ein neues Verzeichnis gelegt. Das ist eine Umverpackung, weil es eine neue Implementierung des Dateiverzeichnisses und der Dateistruktur ist, welche die Pack-Funktion bereitstellt. Allerdings muss der interne Name ebenfalls aktualisiert werden, um die Verknüpfung zwischen der ersten und der dritten Datei zu erhalten. Diese Aktualisierung verändert die Inhaltsinformation, was wiederum bedeutet, dass es sich hier außerdem um eine Transformation handelt. Wenn es sich bei dem internen Namen um einen universellen Identifikator handeln würde, müsste er nicht geändert werden. In diesem Fall würde das normierte Rahmenkonzept, welches den universellen Identifikator unterstützt, die Zuordnung enthalten, die zum Speicherort der dritten Datei führt und müsste daher aktualisiert werden. Dieser Ansatz wäre für ein OAIS von Vorteil, weil er es erlaubt, Aktualisierungen zu zentralisieren und einfacher zu verwalten. Die erforderliche Technologie ist jedoch komplexer und es gibt keine allgemein gültige Vereinbarung über die einzusetzende Identifizierungstechnik.

Das letzte Szenario bezeichnet eine irreversible Transformation, die Umverpackung beinhaltet. Sie tritt ein, wenn die Inhaltsinformation sowohl Dateinamen als auch eine Verzeichnisstruktur und zugehörige Dateieigenschaften beinhalten. Die Inhaltsinformation wird dann mit einer anderen Umsetzung des Dateiverzeichnisses und der Dateinamensstruktur auf einen neuen Datenträger migriert. Die neue Verzeichnisstruktur und die neuen Dateinamen unterstützen weniger Dateieigenschaften.

> Die Bits der Inhaltsinformation und der Erhaltungsmetadaten für eine AIC sind als Gruppierung von AIUs definiert, wobei jede AIU aus drei Dateien auf einer CD-ROM zusammen mit ihren Dateinamen, Dateieigenschaften und Verzeichnisnamen besteht. Die Verpackungsinformation besteht aus den Bits, die zur Implementierung der Datei verwendet werden sowie der Verzeichnisstruktur, die den Zugriff auf jede dieser drei Dateiinstanzen gewährleistet, aber sie enthält nicht die aktuellen Datei- und Verzeichnisnamen. Es kann tausende AIU-Exemplare auf einer einzigen CD-ROM geben. Der Transfer dieser AIC auf einen neuen Datenträger, der eine neue Repräsentation für die Datei- und Verzeichnisstruktur einsetzt, die weniger Dateieigenschaften hat, kann sowohl eine irreversible Transformation zur Folge haben als auch eine Umverpackung. Es ist eine Transformation, weil die Inhaltsinformation, die ursprünglich in der Datei- und Verzeichnisstruktur abgelegt waren, zwischen der neuen Datei- und Verzeichnisstruktur und wahrscheinlich auch innerhalb der Dateien selbst umverteilt werden müssen. Es handelt sich um eine irreversible Transformation, wenn eine algorithmische Eins-zu-eins-Abbildung von der entstandenen Datei, den entstandenen Verzeichnisstrukturen und Dateiinhalten auf die Originaldatei und originalen Verzeichnisstrukturen nicht möglich ist. Es handelt sich um eine Umverpackung, weil es eine neue Umsetzung der Verzeichnis- und Dateistruktur gibt, die als Teil der Verpackung bestimmt wurde. Die Praxis, Inhaltsinformation in einen Datei- oder Verzeichnisnamen zu verschlüsseln, erhöht das Risiko des Informationsverlustes, weil

die Weiterentwicklung einer Datenverwaltungs-Umgebung durch die Möglichkeit, Verzeichnis- und Dateinamen je nach Bedarf zu aktualisieren, vereinfacht wird.

5.1.4 Unterscheidung zwischen transformierten AIPs, bearbeiteten AIPs und abgeleiteten AIPs

Sofern eine digitale Migration keine Transformation mit einschließt, wird davon ausgegangen, dass kein transformiertes AIP erzeugt wurde und es ist nicht erforderlich, dass die Erhaltungsmetadaten des AIPs aktualisiert werden. Mit anderen Worten: Es wird angenommen, dass das transformierte AIP unabhängig ist von Auffrischung, Replikation und Umverpackung, die jeweils keine Auswirkung auf die Inhaltsinformation oder die Erhaltungsmetadaten haben. Das bedeutet nicht, dass das OAIS solche Migrationen nicht nachvollzieht; vielmehr ist es nicht nötig, die Erhaltungsmetadaten als Teil dieser Dokumentation zu aktualisieren. Es wird angenommen, dass das OAIS nachweisen wird, dass solche Migrationen die Inhaltsinformation oder Erhaltungsmetadaten nicht verändert haben und dass jegliche Umverpackung immer noch dieselben Komponenten mit denselben Beziehungen zusammenhält. Es wird ebenfalls angenommen, dass das OAIS das Vorkommen dieser Ereignisse, einschließlich der Verifizierungen, die durchgeführt wurden, als Bestandteil seiner größeren operativen Provenienzsicherung nachvollzieht, da ihm dies einen zusätzlichen Nachweis hinsichtlich der Authentizität seiner Bestände verleiht. Wenn solche Migrationsprozesse vollständig innerhalb des Archivspeichers durchgeführt werden, bleibt die AIP-ID dieselbe und es entstehen keine Auswirkungen auf die Begleitbeschreibung oder Zugriffshilfen.

Eine digitale Migration, die Transformation mit einschließt, hat ein transformiertes AIP wie in 5.1.3.4 definiert zur Folge. In diesem Fall müssen die Erhaltungsmetadaten aktualisiert werden, um das Quell-AIP und seine Version ermitteln zu können und um zu beschreiben, was gemacht wurde und warum. Das neue AIP wird als Ersatz für das Quell-AIP, in dem die Information in höchstmöglichen Umfang bewahrt wurde, angesehen. Das AIP ist ebenfalls neu und die Begleitbeschreibung muss aktualisiert werden. Das besagt nicht, dass irgendwelche Änderungen an den Zugriffshilfen nötig sind, es sein denn, sie sind mit „fest eingebauten“ AIP-IDs umgesetzt worden.

Ein AIP kann in einigen Umgebungen mit der Zeit erweitert oder verbessert werden. Dabei handelt es sich insofern nicht um eine digitale Migration, als das Ziel hier nicht die Erhaltung, sondern eine Erweiterung oder Verbesserung ist. Diese Art der Veränderung eines AIPs wird als Erzeugung eines **bearbeiteten AIPs** bezeichnet. Das neue bearbeitete AIP wird als Ersatz für das vorherige bearbeitete AIP betrachtet, aber es kann ein historisches Interesse daran bestehen, das vorherige bearbeitete AIP aufzubewahren. Das führt ebenfalls zu einer neuen AIP-ID mit denselben Auswirkungen auf die Begleitbeschreibung und die Zugriffshilfen wie eine digitale Migration vom Typ Transformation.

Für ein OAIS könnte es außerdem nützlich sein, ein AIP zur Verfügung zu stellen, das aus einem bestehenden AIP abgeleitet wurde. Das kann durch das Extrahieren einiger Informationen geschehen oder durch die Sammlung von Informationen aus mehreren AIPs, um die Endnutzer besser bedienen zu können. Dieser Typ von AIP wird als **abgeleitetes AIP** bezeichnet. Es ersetzt keines der AIPs, aus denen es abgeleitet wurde und es ist auch nicht das Ergebnis einer digitalen Migration. Es hat ebenfalls eine neue AIP-ID zur Folge und eine neue Begleitbeschreibung. Ebenso können Aktualisierungen an oder sogar neue Zugriffshilfen nötig werden, je nach Art der Implementierung.

5.2 Erhaltung des Zugriffs- und Nutzungsservice

Ein OAIS möchte vielleicht angesichts der sich verändernden Technologie eine bestimmte Art des Zugriffsservice für Endnutzer erhalten. Um einige Problemstellungen bei der Erhaltung des Zugriffsservice zu skizzieren und Begrifflichkeiten zu vermitteln, bietet dieser Unterabschnitt eine Reihe von Anwendungsszenarien.

5.2.1 Auslieferungs-Programmierschnittstelle

Das erste Szenario geht davon aus, dass die vorgesehene Zielgruppe sich wünscht, Anwendungen zu entwickeln, die auf AIPs unter Verwendung einer Programmierschnittstelle (API) zugreifen können, die das OAIS als Zugriffsprogramm unterhält. Möglicherweise entscheidet sich das OAIS sogar dazu, diese API als Alternative zur Produktion und Lieferung eines physischen DIPs für die Auslieferung bereitzustellen. Diese Art von Dienstleistung erlaubt es dem Endnutzer, als Auftraggeber Anwendungen zu entwickeln, die direkten Zugriff auf die AIPs gewähren. Diese Art von Zugriff kann z.B. für gezieltes „Data Mining“ sinnvoll sein, bei der die Erzeugung und der Versand von DIPs, die große AICs umfassen, unpraktisch ist. Die API könnte es einer Anwendung erlauben, virtuell durch eine AIC zu navigieren, die Bits des Inhaltsdatenobjekts der ausgewählten AIU an die Anwendung zu übergeben und die Speicherorte zur Beschaffung der zugehörigen Repräsentationsinformationen und Erhaltungsmetadaten zu ermitteln. Wenn sich die Technologie weiterentwickelt, steigt das OAIS allerdings um auf neue Hardware, neue Datenträger und neue Betriebssysteme. Wenn das OAIS dann für seine Endnutzer weiterhin dieselbe API behalten möchte, muss es eine „Hülle“ um Teile seiner neuen Infrastruktur bereitstellen, um seine Dienstleistungen der bestehenden Schnittstelle zuzuordnen. Die API muss ausreichend dokumentiert und getestet werden, um sicherzustellen, dass sie unter Verwendung des neuen Zugriffsprogramms die Inhaltsinformation der AIU korrekt übermittelt. Dieses Verfahren sollte keine Änderungen an von der Gemeinschaft der Endnutzer entwickelter Software zur Folge haben. Wenn die API für einen Großteil der AIUs des OAIS anwendbar ist oder es eine beträchtliche Anzahl von Endnutzeranwendungen für die API gibt, ist der Einsatz einer „Hülle“ eindeutig plausibel und dürfte in einem vorteilhaften Preis-Leistungs-Verhältnis für das OAIS und seine vorgesehene Zielgruppe stehen. Das „Schichtenmodell der Information“ in Anhang D beschreibt einige potenzielle Standard-APIs näher.

5.2.2 Erhaltung des Look and Feel von Zugriffsprogrammen

Das zweite Szenario geht davon aus, dass die vorgesehene Zielgruppe sich wünscht, dass das ursprüngliche Look and Feel der Inhaltsinformation einer Reihe von AIUs erhalten bleibt, genauso wie es von einer Anwendung oder einer Reihe von Anwendungen angeboten wird. Konzeptionell bietet das OAIS eine Umgebung, die es dem Endnutzer erlaubt, sich die Inhaltsinformation der AIU von den Transformations- und Präsentations-Möglichkeiten der Anwendung anzeigen zu lassen. Es könnte zum Beispiel den Wunsch geben, eine bestimmte Anwendung einzusetzen, die Daten von einer ISO 9660 CD-ROM extrahiert und diese als Multispektralbild darstellt. Diese Anwendung läuft auf einem bestimmten Betriebssystem, erfordert eine Reihe von Steuerungsinformationen, braucht Zugriff auf ein CD-ROM-Lesegerät und stellt die Information über Treibersoftware auf einem bestimmten Anzeigegerät dar. In manchen Fällen könnte diese Anwendung so weit verbreitet sein, dass alle Mitglieder der vorgesehenen Zielgruppe Zugriff auf die Umgebung haben und das OAIS das Inhaltsdatenobjekt lediglich als Bitfolge bestimmt, der von der Anwendung gebraucht wird. Oder aber ein OAIS stellt eine solche Umgebung inklusive Zugriffsprogramm zur Verfügung, wenn die Umgebung weniger leicht zugänglich ist. Mit dem Umzug des OAIS und/oder der vorgesehenen Zielgruppe auf neue Computerumgebungen jedoch wird gleichzeitig die Anwendung gar nicht mehr oder nur noch eingeschränkt funktionieren.

Die Vorgehensweise des OAIS bei der Erhaltung eines Zugriffsprogramms wird wohl zu einem großen Teil davon abhängen, ob es den Quellcode für das Zugriffsprogramm hat oder nicht oder ob es ihn bekommen kann. Unterabschnitt 5.2.2.1 bespricht bewährte Methoden zur Erhaltung des Zugriffs auf Anwendungen über Technologieveränderungen hinweg. Die ausschlaggebenden Faktoren bei der Anwendung dieser Methoden sind das Preis-Leistungs-Verhältnis für das OAIS und die vorgesehene Zielgruppe. Gäbe es einen dringenden Bedarf für das OAIS, das Look and Feel des Zugriffs zu erhalten, der Quellcode oder kommerzielle Brücken wäre aber nicht verfügbar, dann müsste das OAIS mit „Emulations“-Tech-

nologie experimentieren, die zurzeit im Umfeld digitaler Bibliotheken erforscht wird. Diese Technologie wird in 5.2.2.2 besprochen.

5.2.2.1 Methoden mit Verfügbarkeit des Quellcodes

Die Reaktion des OAIS auf die Erhaltung eines Zugriffsprogramms wird zumindest teilweise davon abhängen, ob es den Quellcode für die Anwendung hat oder nicht. Wenn das OAIS über den Quellcode und eine entsprechende Dokumentation der Anwendung verfügt, wird es voraussichtlich versuchen, die Anwendung auf die neue Umgebung zu übertragen und sie ausreichend zu testen, um sicherzustellen, dass sie richtig funktioniert. Wie in 4.2 beschrieben, ist es nicht gerade einleuchtend, wenn die Anwendung zwar läuft, aber falsch funktioniert. Im Idealfall wurden alle möglichen Ausgabewerte zunächst verzeichnet, so dass sie als Basis für die Korrektheitsprüfung des Anschlusses dienen können. Dieser Grad der Überprüfung wird jedoch wahrscheinlich zu einem inakzeptablen Preis-Leistungs-Verhältnis für das OAIS führen. Wenn man bedenkt, dass die Anwendung aus dem originalen Quellcode zusammengesetzt wurde, sind die Algorithmen vermutlich korrekt; die Herstellung einer Teststellung oder die Wiederverwendung einer Teststellung, die mit der Entwurfsdokumentation geliefert wurde, erscheint zweckdienlich.

Falls es sich bei dem Zugriffsprogramm um ein geschütztes, firmeneigenes Paket handelte, das weitverbreitet und kommerziell verfügbar war, gibt es vermutlich eine kommerziell bereitgestellte Brückensoftware (z.B. zur Konvertierung), welche die aktuellen Inhaltsdatenobjekte in andere Formen umwandelt, die vom neuen Zugriffsprogramm unterstützt werden und ein ähnliches Look and Feel vermitteln. Wenn es keine kommerzielle Variante gibt, könnte das OAIS einen Vertrag mit dem Eigentümer des Original-Zugriffsprogramms schließen zur Entwicklung und Bereitstellung eines Quellcodes für ein einfaches Hilfsmittel, das Exemplare von in diesem Format geschriebenen Daten lesen, aber nicht verändern kann. Dieser Ansatz mag aus Kostengründen oder rechtlichen Fragen nicht umsetzbar sein. In allen Fällen wird das OAIS Mechanismen einrichten müssen, um nachzuweisen, dass keine zu erhaltende Information verloren gegangen ist. Das setzt voraus, dass im Vorfeld Kriterien festgelegt worden sind, die genau definieren, was die Inhaltsinformation ausmacht wie in 6 besprochen[20]. Des Weiteren muss das OAIS die Sachverhalte untersuchen, die mit der Gewährleistung der Verfügbarkeit des neuen Zugriffsprogramms für die vorgesehene Zielgruppe verbunden sind.

5.2.2.2 Mögliche Ansätze zur Emulation

Es kann eine zwingende Anforderung aus der vorgesehenen Zielgruppe geben, das Look and Feel eines proprietären Zugriffsprogramms zu erhalten aufgrund der hohen Zahl von AIUs, die von diesem Zugriffsprogramm abhängen. In diesem Fall und wenn das OAIS nicht in der Lage ist, den Quellcode zu bekommen, oder es den Source Code hat, aber nicht die Fähigkeit, die benötigte Anwendung herzustellen, z.B. weil ein Übersetzungsprogramm oder eine Betriebsumgebung nicht verfügbar ist, könnte es unumgänglich sein, den Einsatz einer Emulation zu untersuchen.

Das OAIS könnte erwägen, die Anwendung nachzubilden. Wenn die Anwendung einen gut bekannten Satz an Operationen bereitstellt und eine wohl definierte API für den Zugriff, sollte die API ausreichend dokumentiert und getestet sein, um eine Emulation der Anwendung zu versuchen. Wenn aber die Schnittstelle für Endnutzer in erster Linie ein Anzeige- oder ein anderes Gerät ist, das menschliche Sinne anspricht (z.B. Klang), ist eine Rekonstruktion nahezu unmöglich. Wie in 4.2 beschrieben ist es nicht gerade einleuchtend, wenn die Anwendung zwar läuft, aber falsch funktioniert. Um zu garantieren, dass eine solche Situa-

[20] Vermutlich eine falsche Verweisung im englischen Original, denn Inhaltsinformation wird in den Abschnitten 2.2.2 und 4.2.1.4.1 erläutert.

tion entdeckt würde, müsste man notwendigerweise den Ausstoß des richtig funktionierenden Zugriffsprogramms als Daten mitsamt den entsprechenden Repräsentationsinformationen und Erhaltungsmetadaten protokollieren, so dass diese Information erhalten werden kann. Diese Information muss mit den abgerufenen Ergebnissen nach dem Umzug in eine neue Umgebung abgeglichen werden. Das kann schwierig werden, wenn die Anwendung viele verschiedene Betriebsarten hat. Wenn darüber hinaus die Ausgabe der Anwendung z.B. hauptsächlich an ein Anzeigegerät übermittelt wird, garantiert die Aufzeichnung des Datenstroms nicht, dass die Anzeige in der neuen Umgebung gleich aussieht. Die Kombination von Anwendung und Umgebung muss also dem Endnutzer nicht notwendigerweise vollständig korrekte Informationen geben. Die Pflege eines durchgängigen Look and Feel erfordert als Ausgangspunkt die Erfassung dieses Look and Feel in einer separaten Aufzeichnung, um sie zur Gültigkeitsprüfung heranziehen zu können. Im Allgemeinen dürfte das Look and Feel schwierig, wenn nicht unmöglich, explizit zu spezifizieren sein. Allerdings dürften eine Reihe transformationeller Informationseigenschaften im Wesentlichen Kriterien definieren, gegen welche die Erhaltung getestet werden könnte; Validierung gegen diese Informationseigenschaften würde eine notwendige, wenngleich vielleicht nicht ausreichende Bedingung sein, um die Angemessenheit der Erhaltungsaktivität zu akzeptieren.
Ein anderer Ansatz ist die Emulation der zugrunde liegenden Hardware. Ein Vorteil der Hardware-Emulation besteht in der Annahme, dass sobald eine Hardware-Plattform einmal erfolgreich emuliert wurde, alle Betriebssysteme und Anwendungen, die auf der Original-Plattform liefen, ohne Veränderung auf der neuen Plattform laufen. Allerdings ist hier der Grad der Emulation relevant, z.B. ob sie so weit geht, den Takt der CPU-Befehlsausführung zu duplizieren. Darüber hinaus werden Abhängigkeiten zu Ein-/Ausgabegeräten nicht berücksichtigt. Emulation wurde erfolgreich eingesetzt, um zum Beispiel ein sehr gängiges Betriebssystem auf einem Hardware-System laufen zu lassen, für das es ursprünglich nicht entwickelt wurde, wie z.B. Windows auf einem SUN™ Gerät. Aber auch in einem Fall wie diesem, wenn starke Marktkräfte diesen Ansatz forcieren, müssen nicht alle Anwendungen in der neuen Hardware-Umgebung zwangsläufig richtig laufen oder ausgeführt werden. Es könnte z.B. aufgrund von Einschränkungen in der neuen Hardware-Umgebung vielleicht nicht möglich sein, alle Abhängigkeiten und Taktungen der alten Hardware in vollem Umfang zu simulieren. Wenn die Anwendung einem Menschen Informationen präsentiert, ist es überdies problematisch davon auszugehen, dass irgendwelche neuen Geräte die Information weiterhin richtig präsentieren. Daher ist es naheliegend, eine separate Aufzeichnung über die Präsentation der Information für die Gültigkeitsprüfung anzufertigen. Wenn Emulation erst einmal angewendet worden ist, ist das hierdurch entstandene System besonders anfällig gegenüber zuvor unbekannten Softwarefehlern, die einen kontinuierlichen Zugriff auf die Information ernsthaft gefährden können. Angesichts dieser Einschränkungen scheinen die technischen und wirtschaftlichen Hürden einer Hardware-Emulation beträchtlich zu sein, außer wo die Emulation einem Wiedergabeprozess entspricht, so wie die Anzeige eines Bildes einer Dokumentenseite oder das Abspielen eines Klanges in einem einzigen System.

Es gab Untersuchungen zu alternativen Vorgehensweisen bei der Emulation, wie etwa die Entwicklung einer virtuellen Rechner-Architektur oder Emulation auf der Ebene des Betriebssystems. Diese Verfahren lösen zwar einige Probleme der Hardware-Emulation, werfen aber zugleich neue Fragen auf. Keine der Emulationstechnologien ist für eine fundierte Stellungnahme ausgereift genug. Außerdem drehen sich die laufenden Forschungsprojekte zur Emulation um eine zentralisierte Struktur mit Kontrolle über alle Peripheriegeräte. Der Komplexitätsgrad von Schnittstellen und Dialogeingriffen in einer ubiquitär verteilten Rechner-Umgebung (z.B. www und Java oder allgemeinere Client-Server-Architekturen) und heterogenen Clients könnte aber neue Anforderungen aufwerfen, die über den Anwendungsbereich der derzeitigen Emulationsversuche hinaus gehen.

6 Interoperabilität von digitalen Archiven

Benutzer mehrerer OAIS-Archive können Grund dazu haben, sich etwas Einheitlichkeit oder Kooperation zwischen ihnen zu wünschen.

Endnutzer von mehreren Archiven können sich zum Beispiel wünschen:
- gemeinsame Findmittel zur Unterstützung beim Auffinden von Information über mehrere Archive hinweg;
- ein gemeinsames Paketbeschreibungsschema für den Zugriff;
- ein gemeinsames DIP-Schema für die Auslieferung; oder
- eine einzige umfassende Zugriffsseite.

Produzenten können sich wünschen:
- ein gemeinsames SIP-Schema für die Übergabe an verschiedene Archive; oder
- einen einzigen Aufbewahrungsort für all ihre Produkte.

Manager können sich Mittel wünschen:
- zur Kostenreduzierung durch gemeinsame Nutzung von teurer Hardware und Software und durch gemeinsame Erhaltungsmaßnahmen; und
- zu erhöhter Einheitlichkeit und Qualität der Interaktionen zwischen verschiedenen Archiven.

Es kann daher für Archive von Vorteil sein zu kooperieren, um diese Wünsche zu erfüllen. Die Motivation kann von den Archiven selbst kommen, oder eine Autorität, die Einfluss auf die Archive nehmen kann, ordnet die Kooperation an. Im ersten Fall könnte das Archiv motiviert sein durch:
- den Wunsch, Kosten zu reduzieren;
- den Wunsch, Endnutzer mit ihren Produkten zufrieden zu stellen;
- den Wunsch, Benutzer mit der Qualität ihrer Dienste zufrieden zu stellen; oder
- die Notwendigkeit des Wettbewerbs mit anderen Archiven, um zu überleben oder zu wachsen.

Umstände wie diese können Vereinbarungen anregen und haben dies bereits getan, bei denen kein Bedürfnis nach einem ausdrücklichen Verbund bestand, der als externe Autorität auftritt. Wenn jedoch ein ausdrücklicher Verbund begründet wird, wird die externe Autorität in diesem Referenzmodell vom Management repräsentiert.

Der Zweck dieses Abschnitts ist es, den Grad der Interaktion und Kooperation zwischen den Archiven zu untersuchen. Unterabschnitt 6.1 beschäftigt sich mit den technischen Ebenen der Interaktion, während Unterabschnitt 6.2 Managementfragen, die das Spannungsverhältnis zwischen Kooperation und Autonomie betreffen, diskutiert.

6.1 Technische Ebenen der Kooperation von Archiven

Im Allgemeinen ist ein OAIS mit einem anderen nicht interoperabel; es gibt jedoch eine Reihe von Gründen dafür, dass ein gewisser Grad an Interoperabilität wünschenswert sein kann. Dies kann von Benutzern, Produzenten oder dem Management gewünscht sein. Auch interoperable Archive können unterschiedliche vorgesehene Zielgruppen haben - sogar für die gleichen digitalen Objekte - und folglich unterschiedliche Anforderungen an die Repräsentationsinformation und/oder an die Erschließungsinformationen.

OAIS-Vereinigungen können technisch sowohl durch externe wie auch durch interne Faktoren kategorisiert werden. Externe Faktoren schließen Besonderheiten von Produzentengruppen und Endnutzergruppen mit ein. Interne Faktoren könnten die gemeinsame Implementierung der Informationsmodelle sein, die in 4.2 präsentiert wurden, oder die gemeinsame Nut-

zung eines oder mehrerer Funktionsbereiche, die in 4.1 vorgestellt wurden, durch mehrere Archive.
Dieser Unterabschnitt definiert vier Kategorien von Archivzusammenschluss. Die ersten drei Kategorien sind gekennzeichnet durch sukzessiv höhere Grade der Interaktion:

> *Unabhängig*: Archive, die durch lokale Belange motiviert werden; es gibt keine organisatorische oder technische Interaktion zwischen diesen Archiven.
> *Kooperierend:* Archive mit potenziell gemeinsamen Produzenten, gemeinsamen Übergabestandards und gemeinsamen Auslieferungsstandards, aber ohne gemeinsame Findmittel.
> *Verbunden*: Archive sowohl mit lokaler Nutzergruppe (d.h. der ursprünglichen vorgesehenen Zielgruppe, die vom Archiv bedient wird), als auch einer allgemeinen Nutzergruppe (d.h. einer erweiterten vorgesehenen Zielgruppe), die Interesse am Bestand verschiedener OAIS-Archive hat und diese Archive dazu gebracht hat, den Zugriff auf ihre Bestände mittels eines oder mehrerer gemeinsamer Findmittel zu ermöglichen. Die Zugriffsbedürfnisse der lokalen Nutzergruppe haben jedoch normalerweise Priorität gegenüber denen der allgemeinen Nutzergruppe. Allgemeine Auslieferung und Übernahme sind optionale Funktionen.
> *Gemeinsame Ressourcen*: Archive, die mit anderen Archiven Vereinbarungen zur gemeinsamen Nutzung von Ressourcen getroffen haben, etwa um Kosten zu reduzieren. Das erfordert verschiedene interne Standards in den Archiven (wie z.B. standardisierte Schnittstellen zwischen Übernahme und Speicherung sowie zwischen Zugriff und Speicherung), aber es ändert nicht die Sicht, die die Benutzergruppe auf das Archiv hat.

Der Rest dieses Unterabschnitts wirft einen detaillierteren Blick auf diese Zusammenschlusskategorien.

6.1.1 Unabhängige Archive

Es wird angenommen, dass ein unabhängiges Archiv nur eine vorgesehene Zielgruppe bedient. Das Archiv und die vorgesehene Zielgruppe müssen sich einig werden über die Gestaltung der SIPs, DIPs und der Findmittel. Ein unabhängiges Archiv kann diesen Strukturen formale oder de-facto-Standards zugrundelegen, was eine Kooperation mit anderen Archiven, die dieselben Standards umsetzen, erlauben würde. Allerdings basieren die Gestaltungsentscheidungen, diese Standards zu verwenden, nicht auf der Möglichkeit zur Zusammenarbeit mit anderen Archiven, sondern vielmehr auf lokalen Anforderungen und Kosteneinsparungen.

Die Einstufung eines Archivs als unabhängig basiert nicht auf der Größe des Archivs oder auf verteilter Funktionalität. Ein unabhängiges Archiv kann einen Standort besitzen oder kann über mehrere Standorte verteilt sein. Es kann viele Standards für ein bestimmtes internes Element nutzen. Wenn es jedoch keine Interaktion mit anderen Archiven gibt, dann ist das Archiv unabhängig.

6.1.2 Kooperierende Archive

Kooperierende Archive gründen sich auf Vereinbarungen zwischen zwei oder mehr Archiven über den Einsatz gemeinsamer Standards. Die einfachste Form von Kooperation zwischen Archiven ist, wenn ein Archiv als Endnutzer von Material eines anderen Archivs auftritt. In diesem Fall muss das benutzende Archiv das DIP-Format des produzierenden Archivs als ein SIP-Format unterstützen. Kooperierende Archive haben miteinander in Beziehung stehende Interessensgruppen, so dass sie Daten von anderen kooperierenden Archiven bestellen und übernehmen und möglicherweise gemeinsame Datenproduzenten haben. Gemeinsame Standards für den Zugriff, Übergabe oder Auslieferung von Daten werden nicht vorausgesetzt. Die einzige Anforderung an diese Architektur ist, dass die kooperierenden

Gruppen wenigstens jeweils ein gemeinsames SIP- und DIP-Format für Anfragen zwischen Archiven unterstützen. Der Kontrollmechanismus für diese Art von Operation zwischen Archiven können ereignisbasierte Bestellungen in jedem Archiv sein.

Abbildungen 6-1 und 6-2 illustrieren das Konzept von kooperierenden Archiven

Abbildung 6-1 stellt eine einfache gegenseitige Austauschvereinbarung zwischen Archiven auf einer rudimentären Ebene der Archivinteraktion dar.

Anmerkung: In dieser und den folgenden Abbildungen wird das OAIS als eine mehrfach vernetzte Einrichtung in Anlehnung an den Aufbau von Abbildung 6-1 dargestellt. Der Einfachheit halber wird immer ein Verbund zwischen zwei Archiven aufgezeigt, obwohl das Konzept unendlich erweitert werden kann.

Die wesentliche Voraussetzung für diesen Verbund ist eine Reihe von wechselseitigen Übergabevereinbarungen, ereignisbasierten Bestellungen und Benutzerschnittstellen-Standards, um zu ermöglichen, dass DIPs von einem Archiv als SIPs von einem anderen Archiv übernommen werden. Er setzt voraus, dass eine Art paarweiser Kompatibilität zwischen den Archiven eingeführt wird. Das erfordert nicht zwangsläufig gemeinsame Zugriffs-, Auslieferungs- und Übergabemethoden für alle Beteiligten, auch wenn das den Austausch fördern würde. Dieser Grad von Vereinbarung wäre außerdem nützlich, wenn die Bestände eines Archivs aufgrund von Managemententscheidungen mit denen eines anderen Archivs zusammengelegt oder in ein anderes Archiv überführt würden.

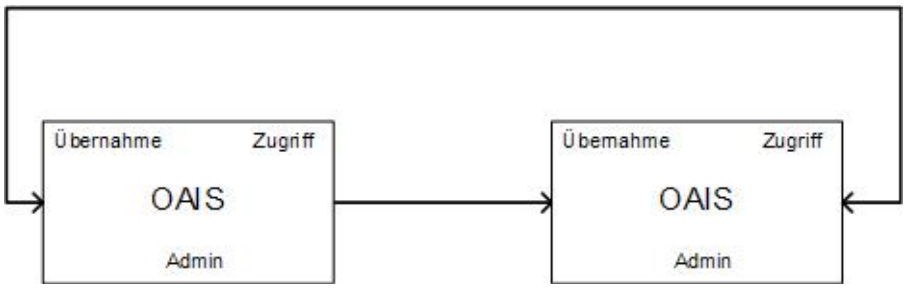

Abbildung 6-1: Kooperierende Archive mit einer gegenseitigen Austauschvereinbarung

Abbildung 6-2 ist ein Beispiel für OAIS-Archive, die ihre Übergabe- und Auslieferungsmethoden zum Vorteil der Benutzer standardisiert haben. Dafür sind keine besonderen externen Elemente notwendig. Der Nachteil liegt darin, dass es keinen formalen Vorgang für den Austausch von Erschließungsinformationen gibt, so dass die Endnutzer separate Recherchesitzungen vornehmen müssen, um für sie interessante AIPs zu finden.

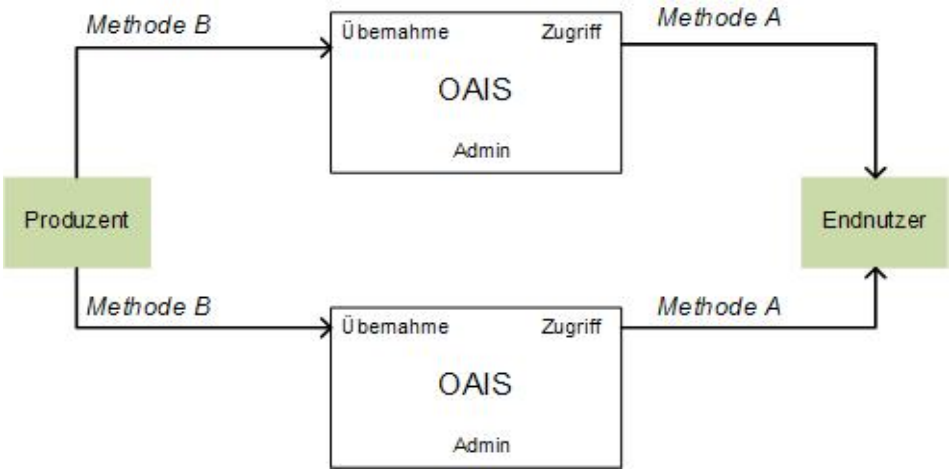

Abbildung 6-2: Kooperierende Archive mit standardisierten Übernahme- und Zugriffsmethoden

6.1.3 Archivverbund

Ein Archivverbund ist seiner Konzeption nach endnutzerorientiert. Zusätzlich zur lokalen Nutzergruppe (d.h. der ursprünglich vorgesehenen Zielgruppe, die von den Archiven bedient wird), existiert eine allgemeine Nutzergruppe (d.h. eine erweiterte vorgesehene Zielgruppe), die Interesse an den Beständen verschiedener OAIS-Archive hat und die diese Archive dazu gebracht hat, den Zugriff auf ihre Bestände über ein oder mehrere gemeinsame Findmittel zu ermöglichen. Der Zugriff für lokale Endnutzer wird jedoch wahrscheinlich Vorrang vor dem für allgemeine Endnutzer haben.

Auf der Verbundebene können externe Elemente eingeführt werden, um die Interoperabilität zu verbessern. Zum Beispiel zeigt Abbildung 6-3 eine funktionelle Architektur, die das Zugriffsproblem auflöst, welches in 6.1.2 beschrieben wurde, indem eine externe Einheit außerhalb der verbundenen OAISe eingeführt wird. Hier haben zwei OAIS-Archive, die ähnliche vorgesehene Zielgruppen haben, einen Verbund geschlossen, um ihren Endnutzern zu ermöglichen, für sie interessante Informationspakete von beiden OAISen mit nur einer Recherchesitzung zu finden. Der gemeinsame Katalog und Manager ist das externe (allgemeine) bindende Element, das als ein gemeinsamer Zugriffspunkt für die Information in beiden Archiven dient. DIPs, die jeweils die Findmittel der beteiligten Archive enthalten, werden in den gemeinsamen Katalog übernommen, wie es die gestrichelten Linien in Abbildung 6-3 zeigen. Der gemeinsame Katalog kann seine Tätigkeit darauf beschränken, als Findmittel zu dienen, oder er kann die gemeinsame Auslieferung von Produkten von einem oder von beiden OAISen einschließen, wie es die gestrichelten Linien in der Abbildung zeigen.

Anmerkung: Der Darstellung werden zusätzliche Zugriffsports hinzugefügt, um die potienzell abweichenden Sichtweisen jeder Endnutzergruppe zu illustrieren.

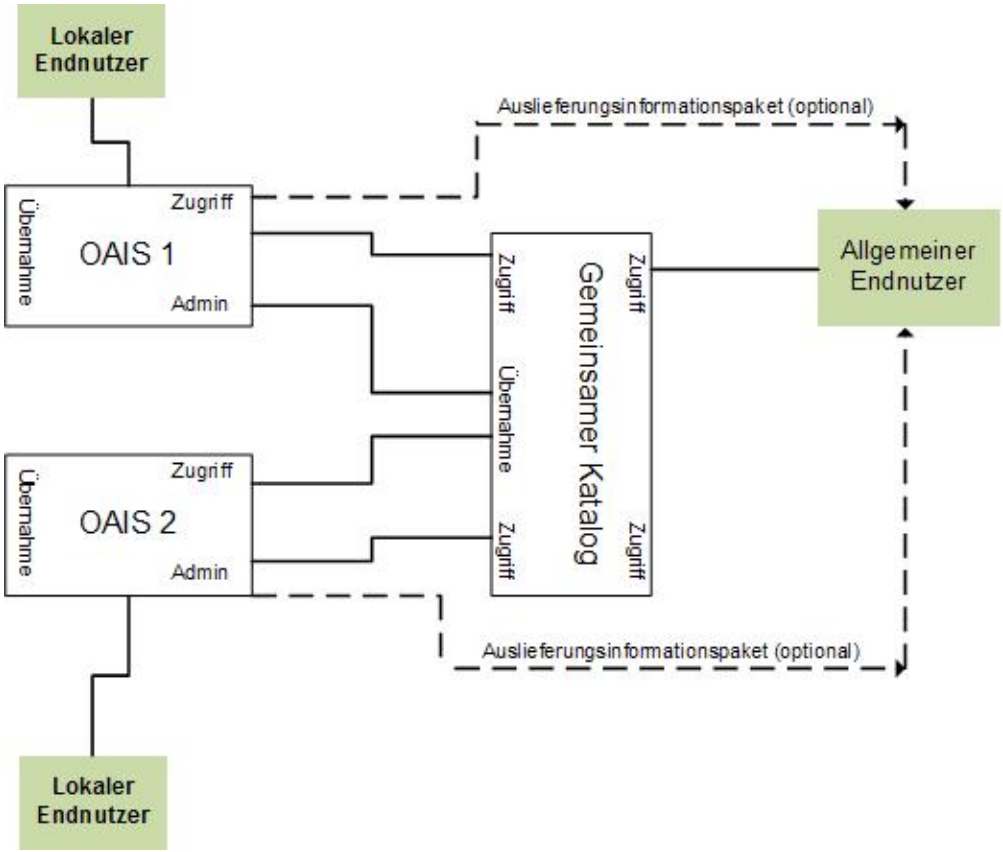

Abbildung 6-3: Ein OAIS-Verbund unter Einsatz eines gemeinsamen Katalogs

Archivverbünde können weiter in drei Funktionsebenen eingeteilt werden.

Zentrale Stelle: Der allgemeine Zugriff wird durch den Export einer Begleitbeschreibung in einem Standardformat zu einer allgemeinen Stelle erreicht. Die allgemeine Stelle verwaltet selbstständig einen Satz von Deskriptoren von vielen Archiven und hat Findmittel, um herauszufinden, welches Archiv eine gesuchte Kollektion besitzt. Dem Endnutzer wird eine kombinierte Ansicht der Bestände von zahlreichen Stellen gegeben, die zentral gepflegt wird. Um Details der Dokumente zu sehen, muss der Benutzer auf die Stelle zugreifen, die das aktuelle Dokument beinhaltet. Das lässt sich einfacher bewerkstelligen, wenn Stellen und Clients einen Standardsatz von Protokollen unterstützen.

Verteilte Findmittel: Der allgemeine Zugriff wird durch einen allgemeinen Knotenpunkt, der eine Anfrage auf mehrere lokale Archive verteilen kann, erreicht. Das bedeutet, dass die lokale Einheit ‚Datenverwaltung' eine zusätzliche Begleitbeschreibung im allgemeinen Format speichern muss oder die allgemeinen Anfragen in lokale Anfrage übersetzt. In diesem Fall wäre eine Möglichkeit, ein gemeinsames DIP-Format zu etablieren, um die Mühe für Endnutzer, die vielleicht Produkte von vielen Archiven bestellt haben, zu erleichtern.

Verteilte Zugriffshilfen: Der allgemeine Zugriff wird dadurch erreicht, dass der Funktionalität der verteilten Findmittel, die oben diskutiert wurde, ein standardisierter Bestell- und Auslieferungsmechanismus hinzugefügt wird; er ist über allgemeine Knotenpunkte zugänglich. Dies ist ein völlig funktionales, verbundenes System. Hier kann das allgemeine System Einfluss auf das Schema der Begleitbeschreibung in jedem lokalen Archiv nehmen; optimal wäre es, neue lokale Archive einzurichten, die auf den allgemeinen Schemata und Findmitteln basieren, um einen hohen Grad an Interoperabilität zu gewährleisten.

Es gibt mehrere wesentliche Fragen, die Policy und Technologie betreffen und die angegangen werden müssen, wenn ein OAIS einem Verbund beitritt oder verschiedene unabhängige OAISe beschließen, einen Verbund zu gründen. Dazu gehören:

Eindeutige AIP-Namen für jedes AIP im Verbund. Ein OAIS hat die Verantwortung, für jedes seiner AIPs eine eindeutige Identität bereitzustellen. Wenn ein OAIS einem Verbund beitritt, gibt es keine Garantie, dass einige von seinen aktuell verwendeten OAIS-AIP-Identifikatoren nicht bereits von anderen Mitgliedern des Verbunds genutzt werden. Ein Beispiel für eine generelle Lösung dieses Problems ist, die AIP-Identifikatoren im Verbund zu erstellen, indem jedem OAIS im Verbund eine eindeutige ID zugewiesen wird und diese ID mit jedem AIP, das von diesem OAIS erhalten wird, verknüpft wird. Dieser OAIS-Name könnte einem Standard entsprechend formatiert sein, welcher dem Kunden oder anderen Verbundsmitgliedern die notwendige Information gibt, um eine Verbindung zu dem Archiv aufzubauen, das die gesuchten AIPs hat. Ein Beispiel für einen Standard, der das leistet, ist die ISO X.500 für Verzeichnisdienste.

AIP-Dubletten in mehreren unterschiedlichen OAISen mit verschiedenen AIP-Namen. Das Problem wird dadurch verursacht, dass vor der Gründung des Verbunds einige OAISe die Inhaltsinformation von AIPs in anderen Archiven duplizierten, um einen lokalen Benutzerzugriff zu ermöglichen. In diesem Fall sieht ein allgemeiner Endnutzer alle Kopien als einzelne, eindeutig bezeichnete AIPs. Eine detaillierte Überprüfung der Erhaltungsmetadaten, die mit der Inhaltsinformation verbunden sind, sollte es dem Endnutzer ermöglichen, die originale, maßgebliche Kopie zu finden, aber der Suchprozess könnte sehr frustrierend für den Benutzer sein. Ein praktikabler Weg damit umzugehen ist, ein Feld in der Begleitbeschreibung aller AIPs anzulegen,

in welchem festgehalten ist, ob es sich um das Original oder eine Kopie handelt. Diese Methode funktioniert jedoch nicht, wenn vor der Gründung des Verbunds zwei oder mehr Archive die Inhaltsinformation von dem Produzenten zur Archivierung erhalten haben. In diesem Fall würden die verbundenen Archive diese AIP-Dubletten als einzigartige, originale AIPs ansehen.

Die Erhaltung des Verbundzugriffs auf AIPs, wenn ein OAIS den Betrieb einstellt. Leider werden viele Archive schließen, während ihre Bestände noch Wert für die Verbundgemeinschaft haben. Der Verbund sollte eine Vereinbarung mit jedem Mitglieds-OAIS haben, die bestimmt, welches OAIS die Verantwortlichkeit für die Erhaltung der Bestände eines geschlossenen OAIS übernimmt.

Benutzer-Authentifizierung und Zugriffsverwaltung für allgemeine Benutzer. Wenn ein OAIS eine Policy hat, die den Zugriff auf einige seiner AIPs einschränkt oder Gebühren für die Auslieferung einiger Informationspakete erhebt, entsteht das Problem, wie der Benutzer, der die Anfragen an den zentralen Knotenpunkt schickt, identifiziert und authentifiziert werden kann. Jedes OAIS wird ein Authentifzierungs- und Zugriffsverwaltungssystem für seine lokalen Benutzer implementiert haben. Die Infrastruktur für diese Funktion muss erweitert werden, um die allgemeinen Benutzer einzuschließen. Hier sind einige Beispiele von Methoden, die in aktuellen Systemen dafür eingesetzt werden:

- Standardprioritäten, bei denen alle Mitglieder des allgemeinen Knotenpunktes einen gemeinsamen Satz von Zugriffsbeschränkungen teilen und der allgemeine Knotenpunkt alle Authentifizierungen übernimmt, um den Endnutzer als rechtmäßigen Benutzer des allgemeinen Knotenpunkts zu überprüfen. In den Mitglieds-OAISen wird bei der Authentifizierung unterstellt, dass alle Anfragen des allgemeinen Knotenpunktes von einem einzelnen Benutzer kommen.

- Ein Berechtigungsnachweisverfahren, wo der spezifische Remote-Benutzer von jedem am Verbund beteiligten Archiv authentifiziert werden kann und der allgemeine Knotenpunkt lediglich als Zwischenstation dient, um den Authentifizierungsdialog zu übermitteln.

Es gibt viele Faktoren, die die Entscheidung beeinflussen, welche dieser Techniken von einem spezifischen Verbund genutzt werden sollte. Das Hauptkriterium ist die Kleinteiligkeit der Zugriffsbeschränkungen im Verbund. Wenn es nur wenige nichtöffentliche Daten gibt und keine Gebühren für die Datenauslieferung erhoben werden, ist eine Policy sinnvoll, die Einschränkungen des Benutzerzugriffs durch die jeweilige Quelle vornehmen lässt, durch die der Benutzer AIPs entdeckt und bestellt. Dafür ist nur eine kleine Modifikation des Authentifizierungssystems des OAIS erforderlich, indem es einfach den allgemeinen Knotenpunkt als Endnutzer hinzufügt. Der allgemeine Knotenpunkt muss Mechanismen bereitstellen, die die allgemeinen Benutzer bei jedem Authentifizierungsmechanismus der verbundenen OAISe identifiziert.

Werden Gebühren für die Auslieferung archivierter Information erhoben oder gibt es erhebliche nichtöffentliche Daten, die eine individuelle Benutzerauthentifizierung erfordern, werden Proxy-Verfahren nicht ausreichen. Benutzerberechtigungsnachweise wie Passwörter und Zertifikate müssen eingesetzt werden. Die Technologien, die diese Unterstützungsmechanismen ermöglichen, befinden sich noch in der Entwicklung.

6.1.4 Archive mit geteilten Funktionsbereichen

In einer Vereinigung von Archiven mit geteilten Funktionsbereichen hat das Management mit Archiven vereinbart, Funktionsbereiche gemeinsam zu nutzen oder zu integrieren. Grund

hierfür kann sein, kostspielige Ressourcen wie etwa hierarchische Datei-Management-Systeme für den Archivspeicher, Peripheriegeräte für die Übernahme oder Auslieferung von Informationspaketen oder Hochleistungsrechner für komplizierte Transformationen zwischen SIPs, AIPs oder DIPs gemeinsam zu nutzen. Diese Vereinigung unterscheidet sich grundsätzlich von den vorangegangenen Beispielen, insofern, dass die interne Struktur des Archivs nicht länger außen vor gelassen werden kann.

Abbildung 6-4 illustriert, wie sich zwei Archive, OAIS 1 und OAIS 2, einen gemeinsamen Speicherbereich, bestehend aus einem Archivspeicher und einer Einheit ‚Datenverwaltung', teilen. Die Zugriffs- und Übernahmeeinrichtungen können auf jedem beliebigen der vorher beschriebenen Grade von Interoperabilität erfolgen. Tatsächlich kann jedes Archiv vollkommen unabhängige Zielgruppen bedienen, wie in dieser Abbildung angedeutet. Allerdings werden Standards für die internen Schnittstellen zwischen Übernahme und Speicher sowie zwischen Zugriff und Speicher gebraucht, damit das gemeinsame Speicherelement erfolgreich arbeiten kann.

Weitere potenzielle geteilte Dienste umfassen Verzeichnisse mit Repräsentationsinformationen und Dienste zur Namensauflösung wie zum Beispiel DNS. Im ersten Fall sollte das Verzeichnis mit Repräsentationsinformationen ebenfalls ein OAIS sein, und die Repräsentationsinformation, die es enthält, sollte Teil der von ihm enthaltenen Inhaltsinformation sein. Die Repräsentationsinformation, die es enthält, könnte beispielsweise Teil des Repräsentationsnetzwerks für die Inhaltsinformation in einem AIP in einem anderen OAIS sein. In diesem Fall kann das OAIS, das dieses AIP enthält, Kopien des Repräsentationsnetzwerks aus dem Verzeichnis cachen. Ob es dies tut oder ob es sich darauf verlässt, dass das Verzeichnis das Repräsentationsnetzwerk pflegen wird: die Verantwortung für die Verstehbarkeit der Inhaltsinformation bleibt letztlich bei dem OAIS, das das AIP enthält.

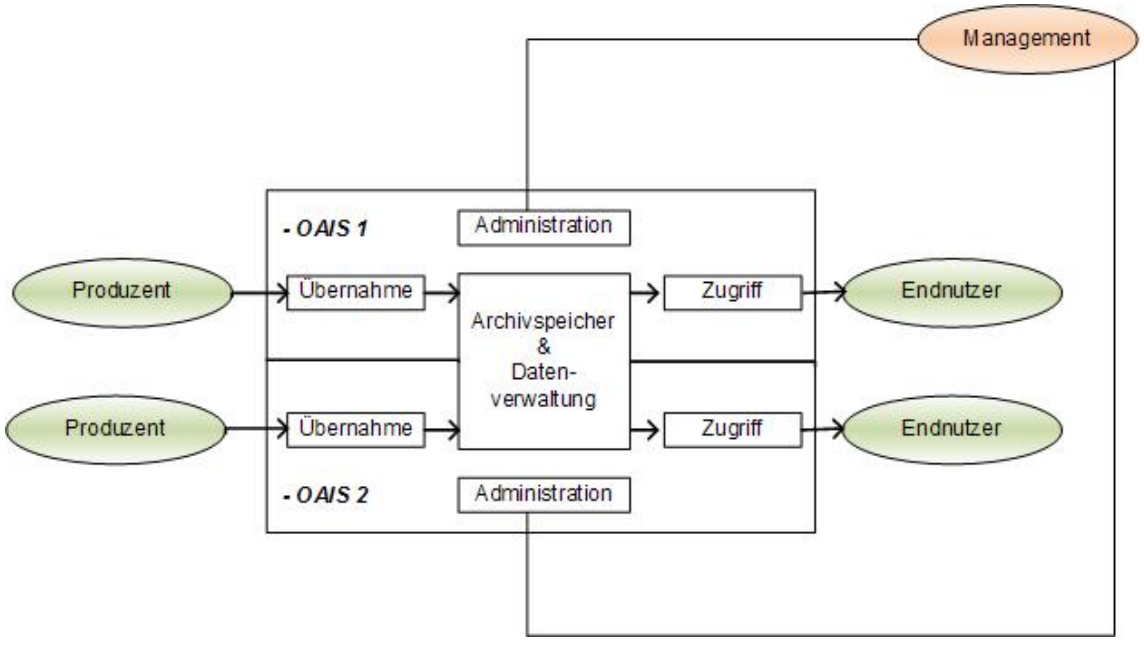

Abbildung 6-4: Archive mit gemeinsam genutztem Speicher

6.2 Managementfragen bei Archivverbünden

Die oben stehenden Beispiele zeigen, dass das OAIS-Modell im Einklang damit steht, bestimmte Ziele im Verbund zu erreichen. Es sollte allerdings berücksichtigt werden, dass einige dieser Ziele durch freiwilliges Handeln erreicht werden können. Das ist eine wichtige Größe bei der Vereinigung von Systemen, einschließlich Archiven, weil sie den Grad der Autonomie für jedes System festlegt. Im Mittelpunkt der Autonomiefrage steht die Leichtig-

keit, mit welcher eine Vereinigung von einem der Beteiligten verändert werden kann. Einige mögliche Charakterisierungen von verschiedenen Graden der Autonomie können sein:

- Keine Interaktionen und daher keine Vereinigung.
- Vereinigungen, die die Autonomie der Mitglieder wahren. Ein Mitglied der Vereinigung muss möglicherweise gewisse Dinge tun, um teilnehmen zu können, kann die Vereinigung aber fristlos und folgenlos verlassen. Ein Beispiel ist die Partizipation im Internet, einschließlich der Tätigkeit als ein Domain-Namen-Server. Ein Mitglied muss gewisse Anforderungen erfüllen, um sich zu beteiligen, einschließlich der Pflege einer Seite mit gewissen Merkmalen, aber das Mitglied wird effektiv von der Vereinigung gestrichen, wenn die Seite sich nicht mehr daran hält. Es gibt jedoch keine Strafe dafür. Daher behalten die Mitglieder die ganze Zeit volle Autonomie, weil sie frei sind straflos zu tun, was sie wollen.
- Vereinigungen, die Vereinsmitglieder vertraglich binden. Um das Wesen dieser Vereinigung zu verändern, muss ein Mitglied den Vertrag neu aushandeln. Wie viel Autonomie erhalten bleibt, hängt davon ab, wie schwierig es ist, die Veränderungen auszuhandeln. Die Schwierigkeit kann steigen, je mehr Beteiligte an dem Vertrag es gibt.

Die Autonomie-Frage ist eine zentrale Frage für interagierende Archive, sie bestimmt die Leichtigkeit, mit der jedes Veränderungen im Wesen der Vereinigung bewirken kann ebenso wie die Auswirkung/Strafe für jedes für die Rückgewinnung der vollen Autonomie. Diese Dimension unterscheidet sich vom Grad der technischen Homogenität, den die Vereinigung implementiert oder unterstützt, sie ist aber nicht völlig unabhängig davon. Zum Beispiel kann in einer breiten Vereinigung, in der es jeder beteiligten Einheit frei steht, die Vereinigung ohne Strafe zu verlassen, ein hoher Grad an technischer Homogenität erreicht werden. Allerdings kann die Dauerhaftigkeit einer solchen Vereinigung nicht garantiert werden, und möglicherweise wird sie verbessert, indem eine Neuaushandlung der Vereinigung erschwert wird oder indem Strafen für die Rückgewinnung der vollen Autonomie eingeführt werden. Außerdem kann ein gegebener Grad technischer Homogenität möglicherweise schneller und mit niedrigeren Kosten erreicht werden, wenn der Vertrag verbindlicher ist.

Anlage zur deutschen Übersetzung

Übersicht über die englischen Glossarbegriffe

Access Functional Entity – Funktionseinheit Zugriff
Access Rights Information – Information über Zugriffsrechte
Access Aid – Zugriffshilfe
Access Collection – Zugriffskollektion
Access Software – Zugriffsprogramm
Adhoc Order – Spontan-Bestellung
Administration Functional Entity – Funktionseinheit Administration
AIP Edition – Bearbeitetes AIP
AIP Version – Transformiertes AIP
Archival Information Collection (AIC) – Archivinformationskollektion (AIC)
Archival Information Package (AIP) – Archivinformationspaket (AIP)
Archival Information Unit (AIU) – Archivinformationseinheit (AIU)
Archival Storage Functional Entity – Funktionseinheit Archivspeicher
Archive – Archiv
Associated Description – Begleitbeschreibung
Authenticity – Authentizität
Collection Description – Kollektionsbeschreibung
Common Services – Allgemeine Dienste
Consumer – Endnutzer
Content Data Object – Inhaltsdatenobjekt
Content Information – Inhaltsinformation
Context Information – Kontextinformation
Co-operating Archives – Kooperierende Archive
Data – Daten
Data Dictionary – Daten-Lexikon
Data Dissemination Session – Datenauslieferungssitzung
Data Management Functional Entity – Funktionseinheit Datenverwaltung
Data Management Data – Daten der Datenverwaltung
Data Object – Datenobjekt
Data Submission Session – Datenübergabesitzung
Derived AIP – Abgeleitetes AIP
Descriptive Information – Erschließungsinformationen
Designated Community – vorgesehene Zielgruppe
Digital Migration – Digitale Migration
Digital Object – Digitales Objekt
Dissemination Information Package (DIP) – Auslieferungsinformationspaket (DIP)
Event Based Order – Ereignisbasierte Bestellung
Federated Archives – Archivverbund
Finding Aid – Findmittel
Fixity Information – Persistenzinformation
Global Community – Allgemeine Nutzergruppe
Independently Understandable – Unmittelbar verstehbar
Information – Information
Information Object – Informationsobjekt
Information Package – Informationspaket
Information Property – Informationseigenschaft
Information Property Description – Informationseigenschaftsbeschreibung
Ingest Functional Entity – Funktionseinheit Übernahme
Knowledge Base – Grundwissen
Local Community – Lokale Nutzergruppe
Long Term – Langfristig

Long Term Preservation – Langzeiterhaltung
Management – Management
Member Description – Teilbeschreibung
Metadata – Metadaten
Non-Reversible Transformation – Irreversible Transformation
Open Archival Information System (OAIS) – Offenes Archiv-Informations-System (OAIS)
Order Agreement – Bestellvereinbarung
Ordering Aid – Bestellsystem
Overview Description – Übersichtsbeschreibung .
Package Description – Paketbeschreibung
Packaging Information – Verpackungsinformation
Physical Object – Physisches Objekt
Preservation Description Information (PDI) – Erhaltungsmetadaten (PDI)
Preservation Planning Functional Entity – Funktionseinheit Erhaltungsplanung
Producer – Produzent
Provenance Information – Provenienzinformation
Reference Information – Referenzinformation
Reference Model – Referenzmodell
Refreshment – Auffrischung
Repacking – Umverpackung .
Replication – Replikation
Representation Information – Repräsentationsinformation
Representation Network – Repräsentationsnetzwerk
Representation Rendering Software – Repräsentations-Wiedergabe-Programm
Retrieval Aid – Bereitstellungssystem
Reversible Transformation – Reversible Transformation
Search Session – Recherchesitzung
Semantic Information – Semantische Information
Structure Information – Strukturinformation
Submission Agreement – Übergabevereinbarung .
Submission Information Package (SIP) – Übergabeinformationspaket (SIP)
Succession Plan – Nachfolgeplan
Transformation – Transformation
Transformational Information Property – transformationelle Informationseigenschaft
Unit Description – Archiveinheits-Beschreibung